The Wiley-Blackwell Handbook of Mood Disorders

Second Edition

Edited by Mick Power
Contents

About the Editor vii
List of Contributors ix
Foreword to First Edition by Kay Redfield Jamison xi

Part I Unipolar Depression 1

1 The Classification and Epidemiology of Unipolar Depression 3
Paul Bebbington

2 Biological Models of Unipolar Depression 39
Anthony J. Cleare and Lena J. Rane

3 Cognitive Models and Issues 69
David A. Grant, Peter J. Bieling, Zindel V. Segal, and Melanie M. Cochrane

4 Psychosocial Models and Issues in Major Depression 87
Antonia Bifulco

5 The Developmental Psychopathology of Depression 107
Katie A. McLaughlin, Louisa C. Michl, and Kate L. Herts

6 Biological Treatment of Mood Disorders 143
Charlotte L. Allan, Anya Topiwala, Klaus P. Ebmeier, David Semple, and Douglas Steele

7 CBT for Depression 173
Mick Power

8 Interpersonal Psychotherapy of Depression 193
John C. Markowitz

9 Marital Therapy for Dealing with Depression 215
Guy Bodenmann and Ashley Randall

10 Depression: The Challenges of an Integrative, Biopsychosocial Evolutionary Approach 229
Paul Gilbert
Contents

Part II  Bipolar Depression  289

11  Classification and Epidemiology of Bipolar Disorder  291

Sameer Jauhar and Jonathan Cavanagh

12  Neurobiological Theories of Bipolar Disorder  311

Karine Macritchie and Douglas Blackwood

13  Psychological Theories of and Therapies for Bipolar Disorder  325

Kim Wright

14  Further Integration of Patient, Provider, and Systems Treatment Approaches in Bipolar Disorder: Where New Evidence Meets Practice Reality  343

Sagar V. Parikh and Sidney H. Kennedy

15  Cognitive Behavioral Therapy for Bipolar Affective Disorders  361

Matthias Schwannauer

16  Self-Management and the “Expert Patient in Bipolar Disorders”  383

Anne Palmer

Part III  General Issues  397

17  Current Approaches to the Assessment of Depression  399

Dave Peck

18  Suicide and Attempted Suicide  413

Andrew K. MacLeod

19  Vulnerability to Depression in Culture, Mind, and Brain  433

Yulia E. Chentsova-Dutton and Andrew G. Ryder

20  Mood Disorders and Chronic Physical Illness  451

Somnath Chatterji and Nicole Bergen

21  Depression in Older People: Cognitive Behavior Therapy, Evidence and Practice  463

Ken Laidlaw

22  Summary and New Directions  485

Mick Power

Author Index  495

Subject Index  523
About the Editor

Mick Power is currently Professor of Clinical Psychology at the Universities of Edinburgh and Tromsø in Norway. For many years he has worked with the World Health Organization to develop a measure of quality of life, the WHOQOL, which is now in widespread use throughout the world. He has written academic books that focus on emotion and the emotional disorders, such as Cognition and Emotion: From Order to Disorder written with his colleague Tim Dalgleish for which the second edition was published in 2008. In 2012 he published a book on psychology and religion, Adieu to God—Why Psychology Leads to Atheism. He is a founding editor of the journal Clinical Psychology and Psychotherapy.
List of Contributors

Charlotte L Allan, Oxford, UK
Paul Bebbington, London, UK
Nicole Bergen, Geneva, Switzerland
Peter J. Bieling, Toronto, Canada
Antonia Bifulco, London, UK
Douglas Blackwood, Edinburgh, UK
Guy Bodenmann, Zurich, Switzerland
Jonathan Cavanagh, Glasgow, UK
Somnath Chatterji, Geneva, Switzerland
Yulia E. Chentsova-Dutton, Washington, USA
Melanie Cochrane, Toronto, Canada
Anthony J. Cleare, London, UK
Klaus Ebmeier, Oxford, UK
David A. Grant, Toronto, Canada
Paul Gilbert, Derby, UK
Kate L. Herts, Boston, USA
Kay Jamison, Columbia, USA
Sameer Jauhar, London, UK
Sidney H. Kennedy, Toronto, Canada
Ken Laidlaw, Edinburgh, UK
Andrew K. MacLeod, London, UK
Karine Macritchie, Edinburgh, UK
John C. Markowitz, New York, USA
Katie McLaughlin, Boston, USA
Louisa C. Michl, Boston, USA
Anne Palmer, Norwich, UK
Sagar Parikh, Toronto, Canada
Dave Peck, Edinburgh, UK
Mick Power, Edinburgh, UK
Ashley Randall, Arizona, USA
Lena J. Rane, London, UK
Andrew Ryder, Concordia, Canada  
Matthias Schwannauer, Edinburgh, UK  
Zindel V. Segal, Toronto, Canada  
David Semple, Edinburgh, UK  
Douglas Steele, Edinburgh, UK  
Anya Topiwala, Oxford, UK  
Kim Wright, Exeter, UK
Moods are so essential to our navigating the world that when they go awry, it is only a matter of time until distress and disaster hit. Moods allow us to gauge people and circumstance, alert us to danger or opportunity, and provide us with the means to convey our emotional and physical states to others. If we act rashly when we ought to be prudent or hang back when we could move forward to advantage, difficulties accrue. Problems compound if, in addition to disruptions in mood, energy, sleep, and thinking also are affected. Disorders of mood which result from this combined disturbance are common, painful, and too often lethal. Fortunately, they are usually treatable.

Scientists and clinicians have learned a remarkable amount about depression and manic depression, or bipolar disorder, during the past decade. These advances in our understanding of diagnosis, pathophysiology, epidemiology, comorbidity, and treatment are lucidly presented in this excellent handbook. There is a strong emphasis upon a complex approach to mood disorders, with the authors providing important coverage of both psychological and biological perspectives on the causes and treatment of depression and mania. The thorny issues of diagnostic categories, the ever-expanding spectrum of pathological into normal affective states, and the unsettled relation of major depression to bipolar disorder, are well addressed, as are the topics of evolutionary psychology, suicide, and pharmacological and psychotherapeutic treatments.

There remain many important questions: ethical and clinical considerations which will arise after the first genes for mood disorders are located; the intriguing psychological issues raised by mania—its relation to violence and creativity, its addictive qualities, and its place in the field of positive emotions; cross-species comparisons which will be possible as a result of mapping the human and other mammalian genomes and the increased understanding of normal moods which will follow from research into more pathological ones. Neuroscientists, clinicians, psychologists, and molecular biologists make a powerful alliance. This handbook gives an outstanding overview of their accomplishments to date and a sense of the excitement to come.

Kay Redfield Jamison, Ph.D.
Professor of Psychiatry
The Johns Hopkins School of Medicine
Part I
Unipolar Depression
The Classification and
Epidemiology of Unipolar
Depression

Paul Bebbington
UCL Mental Health Sciences Unit, UK

Introduction

In the revised version of this chapter, the original structure has largely been maintained. After all, the principles of classification and the problems it poses for the epidemiology of depression remain the same. However, I have summarized the slow progress toward revision of the ICD and DSM classificatory systems. In the original chapter, I provided illustrative examples of the epidemiology of depression in terms of age, gender, life stress, and childhood antecedents of depression. In the current version, I have, somewhat selectively, updated the account of research in these topics, but have changed the emphasis away from well-established issues such as the effect of gender, and toward the epidemiology of depression in the workplace. This seems particularly relevant in a world where inexorable economic advancement is no longer guaranteed and retrenchment has major effects on work and workers.

Classification and Unipolar Depression

Psychiatric classification is quintessentially a medical procedure. The study of medicine is based on the establishment of separate categories of disorder (illnesses, diseases). These are distinguished in terms of particular types of attributes of the people held to be suffering from them. These attributes comprise symptoms (based on self-report) and signs (based on observation). Individual disorders are conceived in terms of concatenations of symptoms and signs, which are termed syndromes. Such syndromes are provisional constructs, whose validity (or otherwise) is then established by using them as the basis of different sorts of theory: etiological theories, and theories of course and

---

1 The terms illness and disease in ordinary usage strongly suggest a biological basis, while disorder is more neutral, and in psychiatry is generally preferred.
outcome, of treatment, and of pathology (Bebbington, 2011; Wing, Mann, Leff, & Nixon, 1978). There is no doubt that the medical approach to malfunction has been a very effective one, generating new knowledge quickly and efficiently by testing out theories of this type (Bebbington, 1997).

Psychiatric disorders are classified in the hope that the classification can provide mutually exclusive categories to which cases can be allocated unambiguously (the process of case identification). The medical discipline of epidemiology is the study of the distribution of diseases (i.e., medical classes) in the population, and is based on categories of this type. It has been a very powerful method for identifying candidate causal factors, and is thus of great interest to psychiatrists and psychologists, as well as to clinicians from other specialties.

The idea of unipolar depression involves the application of this syndromal approach to psychological disturbance, and is therefore primarily a medical concept. However, the distinctiveness of psychiatric syndromes is both variable and limited. Unipolar depression, in particular, resembles, overlaps, and needs distinguishing from other disorders characterized by mood disturbance: anxiety disorders, other depressive conditions, and bipolar mood disorder. The particular problem with bipolar disorder is that its identification depends on the presence of two sorts of episode in which the associated mood is either depressed or predominantly elated. It is distinct from unipolar disorder in a variety of ways (e.g., inheritance, course, and outcome), and the distinction is therefore almost certainly a useful one. However, depressive episodes in bipolar disorder cannot be distinguished symptomatically from those of unipolar depression. As perhaps half of all cases of bipolar disorder start with a depressive episode, this means that unipolar depression is a provisional category—the disorder will be reclassified as bipolar in 5% of cases (Ramana & Bebbington, 1995).

**Symptoms and Syndromes**

The first stage in the establishment of syndromes is the conceptualization of individual symptoms. Symptoms in psychiatry are formulations of aspects of human experience that are held to indicate abnormality. Examples include abnormally depressed mood, impaired concentration, loss of sexual interest, and persistent wakefulness early in the morning. They sometimes conflate what is abnormal for the individual with what is abnormal for the population, but they can generally be defined in terms that are reliable. Signs (which are unreliable and rarely discriminating in psychiatry, and thus tend to be discounted in diagnosis) are the observable concomitants of such experiences, such as observed depressed mood, or behavior that could be interpreted as a response to hallucinations. Particular symptoms (and signs) often coexist in people who are psychologically disturbed, and this encourages the idea that they go together to form recognizable syndromes. The formulation of syndromes is the first stage in the disease approach to medical phenomena, as syndromes can be subjected to investigations that test out the various types of theories described above.

It is often said, in both medical and lay discourse, that psychiatric disorders are like (or just the same as) disorders in physical medicine. This is not strictly true. Self-reports in general medicine relate to bodily sensation and malfunction in a way that
can be linked to pathological processes. Thus the classical progression of symptoms in appendicitis is related straightforwardly to the progression of inflammation from the appendix to the peritoneal lining. Symptoms in psychiatry, in contrast, are essentially based on idiosyncratic mental experiences, with meanings that relate to the social world. Reformulating psychiatric disorders in terms of a supposed biological substrate would therefore result in the conceptualization of a different condition, which would map imperfectly onto the original disorder.

While syndromes are essentially lists of qualifying symptoms and signs, individuals may be classed as having a syndrome while exhibiting only some of the constituent symptoms. Moreover, within a syndrome there may be theoretical and empirical reasons for regarding particular symptoms as having special significance. Other symptoms, however, may be relatively nonspecific, occurring in several syndromes. Even so, clusters of such symptoms may achieve a joint significance. This inequality between symptoms is seen in the syndrome of unipolar depression: depressed mood and anhedonia are usually taken as central, while other symptoms (e.g., fatigue or insomnia) have little significance on their own. This reflects a serious problem with the raw material of human mental experience: it does not lend itself to the establishment of the desired mutually exclusive and jointly exhaustive categories that underpin medical classification.

In an ideal world, all the symptoms making up a syndrome would be discriminating, but this is far from true, and decisions about whether a given subject’s symptom pattern can be classed as lying within a syndrome usually show an element of arbitrariness. The result is that two individuals may both be taken to suffer from unipolar depression despite exhibiting considerable symptomatic differences.

This is tied in with the idea of symptom severity: disorders may be regarded as symptomatically severe either from the sheer number of symptoms or because several symptoms are present in severe degree. In practice, disorders with large numbers of symptoms also tend to have a greater severity of individual symptoms. In classifications that rely on relatively few symptoms to establish diagnoses of depressive conditions, the issue of severity may need to be dealt with by including other markers, particularly impairment of social engagement and activity, and disabilities in self-care.

The Limits of Classification

As classification aspires to “carve nature at the joints,” the empirical relationships between psychiatric symptoms create special difficulties of their own. In particular, symptoms are related nonreflexively: thus, some symptoms are common and others are rare, and in general they are hierarchically related, rather than being associated in a random manner. Rare symptoms often predict the presence of common symptoms, but common symptoms do not predict rare symptoms. Deeply (i.e., “pathologically”) depressed mood is commonly associated with more prevalent symptoms such as tension or worry, while in most instances tension and worry are not associated with depressed mood (Sturt, 1981). Likewise, depressive delusions are almost invariably associated with depressed mood, whereas most people with depressed mood do not have delusions of any kind. The consequence is that the presence of the rarer,
more “powerful” symptoms indicates a case with many other symptoms as well, and therefore a case that is more symptomatically severe. It is because of this set of empirical relationships between symptoms that psychiatric syndromes are themselves quite largely hierarchically arranged. Thus, schizophrenia is very often accompanied by affective symptoms, although these are not officially part of the syndrome. Likewise, psychotic depression is not distinguished from nonpsychotic depression by having a completely different set of symptoms, but by having extra, discriminating symptoms such as depressive delusions and hallucinations.

Leaky Classes and Comorbidity

The operational criteria set up to identify and distinguish so-called common mental disorders cut across the natural hierarchies existing between symptoms. The consequence is that many people who have one of these disorders also meet the criteria for one or more of the others. This comorbidity has generated much interest, and was even incorporated into the titles of recent major US epidemiological surveys (the National Comorbidity Survey and its replication Kessler, McGonagle, Zhoa, et al., 1994, Kessler et al., 2003). Researchers then divide into two camps: those who think the comorbidity represents important relationships between well-validated disorders, and those who think it arises as an artifact of a classificatory system that is conceptually flawed and fails adequately to capture the nature of affective disturbance.

Depression and the Threshold Problem

The final difficulty with the classification of depression is that it involves imposing a categorical distinction on a set of phenomena that look more like the expression of a continuum. The empirical distribution of affective symptoms in the general population is characteristic: many people have a few symptoms, while few people have many.

For some authorities, this pattern of distribution calls into question the utility of a medical classification. It certainly makes case definition and case finding contentious, as decisions have to be made about the threshold below which no disorder should be identified. People who have few symptoms may still be above this threshold if some of their symptoms are particularly discriminating, but in general the threshold is defined by the number of symptoms. There has always been a tendency in medicine to move thresholds down, particularly as many people who may be regarded by primary care physicians as meriting treatment fall below the thresholds of DSM-IV or ICD-10.

The threshold problem has encouraged a considerable literature relating to subthreshold, subclinical, minor, and brief recurrent affective disorders (Schotte & Cooper, 1999). The tendency to extend the threshold downward is apparent in the establishment of the category of dysthymia, a depressive condition characterized only by its mildness (i.e., a lack of symptoms) and its chronicity. It has, nevertheless, become a study in its own right: it has clear links with major depression presumably because it is relatively easy for someone who already has some depressive symptoms to acquire some more and thereby meet criteria for the more severe disorder.
The imposition of a threshold on an apparent continuum would be less arbitrary if it were possible to demonstrate a naturally occurring “step-change” in the distribution. Thus, while the distribution of IQ is largely continuous, Penrose (1963) noted a clear excess of subjects at the bottom of the continuum who are characterized by a distinct and identifiable pathology. Many have argued that no such distinction exists in affective symptoms (Goldberg, 2000; Tyrer, 1985). While it might be possible to create a threshold that represented a step-change in social disability (Hurry, Sturt, Bebbington, & Tennant, 1983), the evidence does, overall, suggest that affective symptoms are distributed more like blood pressure than IQ. Melzer, Tom, Brugha, Fryers, and Meltzer (2002) used symptom data from the British National Survey of Psychiatric Morbidity to test out the smoothness of the distribution. A single exponential curve provided the best fit for the whole population, but there were floor effects that produced deviations at symptom counts from zero to three. Truncation of the data to take account of this provided an excellent fit (Figure 1.1). This was not affected by selecting subgroups characterized by especially high or low prevalence for analysis.

It can be concluded from this discussion that the epidemiological literature on depressive disorder will need to be interpreted cautiously. We have disorders that are identified as classes imposed on what is empirically a continuum, and which in any event overlap each other. This is made worse because the classificatory schemes are changed at regular intervals. Moreover, two major schemes exist side-by-side. Added to this is the issue of how the symptoms of common mental disorders can be elicited, identified, and used in order to decide if together they can be said to constitute a case.
Competing Classifications

The indistinctness of psychiatric syndromes and of the rules for deciding if individual disorders meet symptomatic criteria has major implications for attempts to operationalize psychiatric classifications. Two systems have wide acceptance: the Diagnostic and Statistical Manual (DSM) of the American Psychiatric Association and the World Health Organization’s International Classification of Disease (ICD). In the early days, revision of classificatory schemata relied almost wholly on clinical reflection. However, since the classifications are set up primarily for scientific purposes, they should properly be modified in the light of empirical research that permits definitive statements about their utility. The standardized and operationalized classifications now available offer an opportunity for using research in this way, and current attempts to modify them are being based on extensive reviews of the evidence.

In the past, much of the pressure for change originated in clinical and political demands. In particular, the establishment of a diagnosis is central to accessing health care in the United States, in distinction to the more problem-based approach in Europe. Moreover, revisions sometimes had the appearance of tinkering in order to capture some imagined essence of the disorders included (Birley, 1990). What looks like fine-tuning can nevertheless make considerable differences to whether individual cases meet criteria or not, and thus disproportionately affects the putative frequency of disorders. We should only jettison classifications on grounds of inadequate scientific utility and as seldom as possible, since too rapid revision defeats the objective of comparison.

Like all such classifications, DSM and ICD are created by committees. It can be argued that the natural tendency for horse trading between experts selected precisely because they are powerful and opinionated has led to an overelaborate structure, an excess of allowable classes and subclasses, and complicated defining criteria. Thus, in DSM-IV (American Psychiatric Association [APA], 1994) there are potentially 14 categories to consider before allocating someone with depressed mood, and in ICD-10 (World Health Organization [WHO], 1992b) there are 22. Greater utility would probably accrue from limiting the primary categories to three (bipolar disorder, unipolar depressive psychosis, and unipolar nonpsychotic depression), and epidemiological research often uses these categories in any case.

Box 1.1 provides a comparison of the definitions of depressive disorder under DSM-IV (APA, 1994) and ICD-10 (WHO, 1992b), slightly simplified. Over the years there has been considerable convergence between the systems. Nevertheless, the differences remain important. The categories are too close together for empirical studies to establish their relative validity, as this would demand enormous samples. However, they are far enough apart to cause discrepancies in identification. Relatively severe cases are likely to be classified as a depressive disorder under both systems. However, milder disorders may be cases under one system and not the other. This becomes important in epidemiological studies of depressive disorder in the general population because such studies usually report their results under one system or the other, and the degree of comparability is hard to quantify. Thus, the use of different classificatory systems is one barrier to comparison between studies; there are others.
Box 1.1 Criteria for depressive episode

<table>
<thead>
<tr>
<th>DSM-IIIR/DSM-IV</th>
<th>ICD-10</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Symptoms present nearly every day in the same 2-week period</strong></td>
<td><strong>Episode must have lasted at least 2 weeks with symptoms nearly every day</strong></td>
</tr>
<tr>
<td><strong>Change from normal functioning</strong></td>
<td><strong>Change from normal functioning</strong></td>
</tr>
<tr>
<td><strong>Key symptoms (n = 2)</strong></td>
<td><strong>Key symptoms (n = 3)</strong></td>
</tr>
<tr>
<td>Depressed mood</td>
<td>Depressed mood</td>
</tr>
<tr>
<td>Anhedonia</td>
<td>Anhedonia</td>
</tr>
<tr>
<td><strong>Ancillary symptoms (n = 7)</strong></td>
<td><strong>Fatigue/loss of energy</strong></td>
</tr>
<tr>
<td>Fatigue/loss of energy</td>
<td>Fatigue/loss of energy</td>
</tr>
<tr>
<td>Weight/appetite loss/gain</td>
<td>Weight and appetite change</td>
</tr>
<tr>
<td>Insomnia/hypersomnia</td>
<td>Sleep disturbance</td>
</tr>
<tr>
<td>Observed agitation/retardation</td>
<td>Subjective or objective</td>
</tr>
<tr>
<td>Low self-esteem/guilt</td>
<td>Agitation/retardation</td>
</tr>
<tr>
<td>Impaired thinking/concentration</td>
<td>Low self-esteem/confidence</td>
</tr>
<tr>
<td>Suicidal thoughts</td>
<td>Self-reproach/guilt</td>
</tr>
<tr>
<td><strong>Criteria: 1 key, 5 symptoms in total</strong></td>
<td><strong>Criteria</strong></td>
</tr>
<tr>
<td><strong>Plus</strong></td>
<td><strong>Mild episode: 2 key, 4 symptoms in total</strong></td>
</tr>
<tr>
<td>Significant distress</td>
<td>Moderate episode: 2 key, 6 symptoms in total</td>
</tr>
<tr>
<td><strong>Or</strong></td>
<td><strong>Severe episode: 3 key, 8 symptoms in total</strong></td>
</tr>
<tr>
<td>Social impairment</td>
<td>Exclusions</td>
</tr>
<tr>
<td><strong>Exclusions</strong></td>
<td>No history (ever) of manic symptoms</td>
</tr>
<tr>
<td>Not mixed episode</td>
<td>Not substance related</td>
</tr>
<tr>
<td>Not substance related</td>
<td>Not organic</td>
</tr>
<tr>
<td>Not organic</td>
<td>Not psychotic</td>
</tr>
<tr>
<td>Not bereavement</td>
<td></td>
</tr>
</tbody>
</table>

It is interesting to see the effect of applying algorithms for the diagnostic categories defined by different systems to a common set of symptom data. The Schedules for Clinical Assessment in Neuropsychiatry (SCAN; WHO, 1992a) allow diagnosis under both DSM and ICD. Table 1.1 shows the effect of applying ICD-10 and DSM-IV criteria to the dataset from a community survey (McConnell, McClelland, Gillespie, Bebbington, & Houghton, 2002) on the identification of cases of depressive episode (ICD) and depressive disorder (DSM). Of the 18 participants diagnosed with a depressive condition by either classification, two-thirds were diagnosed by both. Five cases of depressive episode were not diagnosed as DSM depressive disorder, whereas only one case of depressive disorder was not diagnosed as an ICD depressive episode. In contrast, DSM recognized many more cases of anxiety disorder. Fifteen of the cases defined by DSM were not classed as anxiety disorders by ICD, while only two classified...
Table 1.1.
DSM-III-R and ICD-10 Classification Based on the Same Symptom Data: The Derry Survey

<table>
<thead>
<tr>
<th></th>
<th>No depressive diagnosis</th>
<th>Depressive disorder DSM</th>
</tr>
</thead>
<tbody>
<tr>
<td>No depressive diagnosis</td>
<td>289 (94%)</td>
<td>1 (0.3%)</td>
</tr>
<tr>
<td>Depressive episode ICD-10</td>
<td>5 (1.6%)</td>
<td>12 (3.9%)</td>
</tr>
<tr>
<td>Kappa</td>
<td></td>
<td>0.79</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>No anxiety diagnosis</th>
<th>Anxiety disorder DSM</th>
</tr>
</thead>
<tbody>
<tr>
<td>No anxiety diagnosis</td>
<td>269 (87%)</td>
<td>15 (4.9%)</td>
</tr>
<tr>
<td>Anxiety disorder ICD</td>
<td>2 (0.7%)</td>
<td>21 (6.8%)</td>
</tr>
<tr>
<td>Kappa</td>
<td></td>
<td>0.68</td>
</tr>
</tbody>
</table>

by ICD were not so classed by DSM. Thus, the ICD criteria appear to be less stringent for depressive episode, while the reverse is true of anxiety. The results suggest that the difference between the two systems arises because of differing thresholds rather than because of wide differences in the symptom contents of the classes.

Revising the Classifications

It has been planned to publish new versions of both dominant classificatory systems: DSM-V in 2013, and ICD-11 in 2015. Around 20 years will then have passed since the previous revision of each system (Sartorius, 2010). This is a much longer gap than between previous editions, and represents an improvement: science is not well served by too frequent revision, which needlessly obstructs the possibility of comparison between studies. Some of the delay was inevitable, given the increasing complexity of the process of revision. It involves very many stakeholders, and the establishment by WHO and APA, respectively, of taskforces, advisory groups, and subgroups. The obligation to consult widely involves enormous amounts of work and demands complex processes of integration.

Three very appropriate criteria have been set out for the removal or introduction of categories. These relate to public health, practical utility, and empirical evidence (Sartorius, 2010). However, the precise application of the criteria is likely to lead to disagreement. The gathering and evaluation of the relevant empirical evidence is extremely time-consuming. Moreover, despite the rigor of the procedures involved, political considerations may sometimes trump these three criteria. This is not surprising: classification can have consequences for the way particular university departments and disciplines are funded, the licensed use of drugs is related to diagnostic entities, and, in the United States particularly, so is insurance cover for specific treatments.

A harmonization group has been set up, tasked with the work of ensuring that DSM-V and ICD-11 are as closely compatible as possible in their classificatory procedures. First (2009) has provided an impressively exhaustive article (particularly in its online version) on the potential for harmonization. Of the 176 sets of criteria that correspond in the ICD-10 and DSM-IV systems, only one was identical. Twenty-one percent of sets had conceptually based differences, while 78% had deficiencies that
The basis of epidemiology is case identification. The process of diagnosis involves allocating symptom patterns to a diagnostic class according to given rules. In recent years, these rules have been set out explicitly in diagnostic criteria for research (DCRs) serving the dominant classifications of ICD and DSM. These are so precise that they can be incorporated into computer algorithms like CATEGO (Wing et al., 1990) and OPCRIT (McGuffin, Farmer, & Harvey, 1991).

Once the presence of symptoms has been established, the information can be entered into one of these computer programs in order to provide a diagnostic classification. Human idiosyncrasy is reduced to an absolute minimum in this process. However, researchers must still decide how carefully the underlying symptoms should be identified. The choices include unstructured clinical assessment, responses to questionnaires, and semi-structured research interviews.

The first option, unstructured clinical judgment, introduces variability into the process of case allocation, since researchers are relying for consistency merely on their devotion to a common educational tradition. This situation is made worse when the judgments of an unspecified number of others (e.g., the treating physician) are used, as with the diagnostic information recorded in case registers or in national statistics.

In order to be practicable, questionnaires should seek simple responses to unelaborated questions. However, symptoms are traditionally recognized through an assessment of mental experiences, the subtlety of whose formulation demands quite elaborate inquiry (Brugha, Bebbington, & Jenkins, 1999). They are usually established by a process of clinical cross-examination. This is rather complicated, since it requires the questioner to frame further questions in a flexible way in the light of the answers given by the subject. While it might be possible to encapsulate this procedure in a standard questionnaire by using a branching algorithm, it would be exhaustive and exhausting—it might require paths comprising over a dozen questions just to establish, say, the presence of pathologically depressed mood. In these circumstances there are clearly practical limits to the process of standardization, and it is probably better to rely on the shortcuts available from using the skills of trained clinicians. Since diagnosis
is built around symptoms defined and elicited in this manner, redefinition in terms of answers to much more limited questions would involve changing the concept of the diagnosis itself. No one has seriously suggested that the way psychiatric symptoms are conceptualized should be changed, so if a questionnaire is used, phenomena may be recorded as present or absent when subsequent clinical inquiry might reveal otherwise. Nevertheless, structured questionnaires do allow lay interviewers to be used, with considerable cost savings. The Diagnostic Interview Schedule (DIS; Robins, Helzer, Croughan, & Ratcliff, 1981) and the Composite International Diagnostic Interview (CIDI; Robins et al., 1988) are examples of fully structured questionnaires that have been widely used and have good reliability.

Semi-structured research interviews are costly in clinical time, and the way in which symptoms are established makes it impossible to standardize the procedure entirely (Robins, 1995). Because of the reliance on clinical judgment and the effect this has on the choice of follow-up questions, some variability will remain. This is the price paid for greater validity, that is, the closer approximation to the clinical consensus about the nature of given symptoms. The SCAN (Wing et al., 1990) are based on a semi-structured interview, and have been quite widely used in epidemiological research studies (Ayuso-Mateos et al., 2001; Bebbington, Marsden, & Brewin, 1997; McConnell et al., 2002; McManus, Meltzer, Brugha, Bebbington, & Jenkins, 2009; Meltzer, Gill, Petticrew, & Hinds, 1995; Singleton, Bumpstead, O’Brien, Lee, & Meltzer, 2001). SCAN has good inter-rater reliability despite its semi-structured format.

Questionnaires and Interviews

If, as I have argued, there are doubts in principle about the validity of structured questionnaires, it is worth knowing how their performance compares with semi-standardized interviews. One head-to-head comparison has been made between SCAN and CIDI (Brugha, Jenkins, Taub, Meltzer, & Bebbington, 2001). This permits two separate questions: does the questionnaire provide a similar frequency of disorder to that established by the semi-structured interview? And to what extent are the same cases identified by the two instruments? Differences in frequencies would, at the very least, indicate some systematic biases separating the instruments. However, even if, for example, CIDI recognized more cases than SCAN, it could still be the case that CIDI picked up most or all of the cases identified by SCAN. This would imply that the constraints of a rigid questionnaire tended to lower the threshold of case identification, as might be the case if the rigidity, and the paucity of elaborative questions, led to over-recognition of specific symptoms. If on the other hand, in addition to over-recognition of cases, there were little overlap between the cases found by the two systems, it would indicate a more general failure of rigid questioning to establish symptoms properly.

Brugha et al. (2001) found that the coefficients of concordance for the various ICD-10 diagnoses varied between poor and fair. They calculated that using CIDI would give prevalences about 50% greater than those obtained from SCAN. The index of agreement for any depressive episode was poor (0.14). As expected, the discrepancies arose particularly from cases around the threshold for recognition.
However, we must also take into account the fact that the criteria for depressive disorder (DSM-IV) are more restrictive than those of ICD-10 major depressive episode. This should result in lower prevalence, perhaps 20% lower. There are therefore two influences on prevalence, of opposite effect, whose interaction will be responsible for a methodologically based discrepancy in prevalence. Thus, epidemiological studies reporting DSM-IV major depressive disorder often use CIDI, and this combination probably results in prevalences of depression around 20% above the output from a SCAN/ICD-10 combination. The short form of the CIDI, as used in the Finnish study (Lindeman et al., 2000), may result in particularly high prevalence (Patten, 1997, 2000).

The good news is that as most cases in dispute will lie around the threshold, their attributes are likely to be similar and hence the demographic and social characteristics of the disorder in question are likely to be identified with a fair degree of consistency and accuracy.

**Bottom-Up and Top-Down Case Identification**

The other way in which instruments differ is whether they are diagnosis driven or symptom driven. Instruments that are diagnosis driven do not require eliciting the same set of symptoms in each case in order to establish the appropriate diagnostic category. All they have to do is to confirm that the required diagnostic criteria are met. The DIS and CIDI are examples of such instruments. The advantage is that they can cut corners by not having to check out all symptoms once a diagnosis has been made: this is often the way clinicians work in their ordinary practice.

Symptom-driven instruments however are exhaustive in their coverage of symptoms, and only then do they use the symptomatic information to check if diagnostic criteria have been met (e.g., SCAN, CIS-R). This has several advantages. The first is that in theory it should be possible to use the symptom information to serve a new algorithm if the diagnostic criteria were changed. This might be extremely arduous in practice, although attempts of this sort have been made. A further advantage is of particular relevance to the study of the common affective disorders. Establishing whether or not a set range of symptoms is present allows an overall symptom count to be made, and this is useful when it is appropriate to study the distributions of symptoms in the general population, as in the study by Melzer et al. (2002) mentioned above. In principle, it could also be used to identify more severe disorder, without invoking extrinsic attributes like social performance. Finally, the establishment of individual symptoms in epidemiological samples allows them to be studied in their own right as reflections of psychological processes.

**The Frequency of Depressive Disorder**

In their seminal report on the Global Burden of Disease, Murray and Lopez (1996) projected that by 2020 depression would rank as the second leading cause of disability worldwide. This prediction however relies on the assumption that reasonably accurate
statistics are available, and that they can be integrated across jurisdictions. At the time the estimates were made, these requirements had been met only in the most tenuous way. While things have definitely improved, particularly in the past 10 years, it does remain difficult to calculate the burden of depressive illness in different countries. Differences in the frequency of depression between countries may in part be substantive, but will inevitably be clouded by measurement issues. These include local constraints on the conceptualization and acknowledgement of depression, and variation in the performance of instruments in local hands. Effective quality control will reduce, but not eliminate, such methodological “noise.”

In epidemiological studies, frequency can be measured in a variety of ways: incidence; point, period, and lifetime prevalence; and morbid risk. Box 1.2 defines commonly used rates in epidemiology. General population surveys usually report period or lifetime prevalence rates, while investigations of clinical series often use first contact or admission as a proxy for incidence. In this chapter, I shall rely largely on community studies of prevalence, as the characteristics of clinical series are distorted by nosocomial factors, such as the determinants of, and barriers to, referral to services.

The earliest community psychiatric surveys date back a century, but standardized methods of assessment allowing the comparison of research from different locations have been used only in the past 30 years. The earlier studies have been reviewed elsewhere (Bebbington, 1997, 2004; Weissman et al., 1996). The range of values for prevalence was appreciable, and somewhat greater than in more recent surveys. Moreover, the detailed results did not lend themselves to simple explanation, and there were considerable differences in the information gathered, and in the way it was gathered and combined.

Community psychiatric surveys based on standardized instruments were initially carried out in small areas, as this was relatively easy to organize, even when there were large numbers of subjects. The Epidemiologic Catchment Area surveys used the DIS to interview nearly 20,000 subjects, but were restricted to five localities in the United States (Robins & Regier, 1991). The overall lifetime prevalence of major depression was 4.9%, ranging from 3% to 5.9% in the different centers.

However, it is quite difficult to make sense of differences in prevalence in different locations in psychiatric community surveys, unless the geographical coverage is large.

---

**Box 1.2 Epidemiological rates**

- **Incidence rate:** the number of new cases in a given period as a proportion of a population at risk.
- **Point prevalence rate:** the number of cases identified at a point in time as a proportion of a total population.
- **Period prevalence rate:** the number of cases identified as in existence during a specified period as a proportion of a total population.
- **Lifetime prevalence rate:** a variant of period prevalence where the period for case identification comprises the entire lifetime of each subject at the point of ascertainment.
This is one argument for having surveys of representative national populations. The first of these occurred in Britain (Jenkins et al., 1997), and there have now been three British National Surveys of Psychiatry Morbidity (Jenkins et al., 1997; McManus et al., 2009; Singleton et al., 2001), and two in Australia (Andrews, Hall, Teesson, & Henderson, 1999; Henderson, Andrews, & Hall, 2000; Slade, Johnston, Oakley Browne, Andrews, & Whiteford, 2009). They each involved interviews carried out by nonclinical interviewers with several thousand subjects selected at random from the whole national population. The British surveys were all based on the revised version of the Clinical Interview Schedule (CIS-R; Lewis, Pelosi, Araya, & Dunn, 1992), an interview that provides ICD-10 diagnoses (WHO, 1992a), while the Australian surveys used variants of CIDI (Robins et al., 1988). CIDI allows both DSM-IV and ICD-10 diagnoses, and the Australian series reported the latter.

The last decade has seen an ambitious attempt to improve the validity of international comparison. The WHO World Mental Health Initiative sought to minimize methodological noise by adopting a common instrument and common methods. The initiative is enormous: at the last count, surveys were being carried out in 28 countries. Some of these surveys were truly national, while others were of specific regions within a given country. The expectation is that over 154,000 people will eventually have been interviewed. The six European national ESEMeD samples are included in the World Mental Health Survey reports (Alonso et al., 2002).

The sheer number of the constituent surveys leads to inherent problems of interpretation. Frequencies may vary because of differential success in engaging participants, different population structures, and the lexical consequences of differences in the transcultural interpretation of emotion (Bebbington & Cooper, 2007). International variations in the ecological context of the population may be important: economic performance, the level of inequality, the degree of urbanization, and the extent of democratic freedoms. Some jurisdictions will be subject, intermittently or persistently, to the effects of war. These influences, identifiable at the national level, will have idiosyncratic impacts on individuals. Ideally, the analysis of international differences in social and other environmental influences should inform our understanding of the nature of psychiatric disorder, but complex results will often elude easy interpretation. However, if associations are observed consistently across different jurisdictions, this does add strength to any conclusions we might draw.

Bromet et al. (2011) provide a relevant example of an attempt to derive synthetic conclusions from the mass of data that has emerged from the World Mental Health Survey Initiative (see Table 1.2). They integrated the findings regarding the prevalence of major depressive episode in 10 high-income and 8 low- to middle-income countries. (The authors did not include participating surveys from Nigeria and Ethiopia because they were suspicious that the low rates of depression in those countries may have resulted from particular difficulties with the interview.) While some response rates were lower than would be regarded as necessary for a representative sample (46% in France), response rates were unrelated to the reported prevalence of depressive disorder.

*Lifetime prevalence* was significantly greater in the high-income countries, averaging 14.6%, compared to 11.1% in the low- to middle-income countries. In contrast, the range of values seen for *12-month prevalence* was much less, and very similar in the
Table 1.2.
Twelve-month and Lifetime Prevalence Percentage of Major Depressive Episode in 18 Countries: World Mental Health Survey

<table>
<thead>
<tr>
<th>Country</th>
<th>12-month prevalence (%)</th>
<th>Lifetime prevalence (%)</th>
<th>12-month/lifetime (%)</th>
<th>Age of onset (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgium</td>
<td>5.0</td>
<td>14.1</td>
<td>35.2</td>
<td>29.4</td>
</tr>
<tr>
<td>France</td>
<td>5.9</td>
<td>21.0</td>
<td>27.9</td>
<td>28.4</td>
</tr>
<tr>
<td>Germany</td>
<td>3.0</td>
<td>9.9</td>
<td>30.1</td>
<td>27.6</td>
</tr>
<tr>
<td>Israel</td>
<td>6.1</td>
<td>10.2</td>
<td>59.6</td>
<td>25.5</td>
</tr>
<tr>
<td>Italy</td>
<td>3.0</td>
<td>9.9</td>
<td>30.2</td>
<td>27.7</td>
</tr>
<tr>
<td>Japan</td>
<td>2.2</td>
<td>6.6</td>
<td>33.3</td>
<td>30.1</td>
</tr>
<tr>
<td>Netherlands</td>
<td>4.9</td>
<td>17.9</td>
<td>27.3</td>
<td>27.2</td>
</tr>
<tr>
<td>New Zealand</td>
<td>6.6</td>
<td>17.8</td>
<td>37.0</td>
<td>24.2</td>
</tr>
<tr>
<td>Spain</td>
<td>4.0</td>
<td>10.6</td>
<td>37.5</td>
<td>30.0</td>
</tr>
<tr>
<td>US</td>
<td>8.3</td>
<td>19.2</td>
<td>43.1</td>
<td>22.7</td>
</tr>
<tr>
<td>Average</td>
<td>5.5</td>
<td>14.6</td>
<td>37.7</td>
<td>25.7</td>
</tr>
<tr>
<td>Brazil</td>
<td>10.4</td>
<td>18.4</td>
<td>56.7</td>
<td>24.3</td>
</tr>
<tr>
<td>Colombia</td>
<td>6.2</td>
<td>13.3</td>
<td>46.7</td>
<td>23.5</td>
</tr>
<tr>
<td>India</td>
<td>4.5</td>
<td>9.0</td>
<td>50.0</td>
<td>31.9</td>
</tr>
<tr>
<td>Lebanon</td>
<td>5.5</td>
<td>10.9</td>
<td>50.0</td>
<td>23.8</td>
</tr>
<tr>
<td>Mexico</td>
<td>4.0</td>
<td>8.0</td>
<td>50.0</td>
<td>23.5</td>
</tr>
<tr>
<td>China</td>
<td>3.8</td>
<td>6.5</td>
<td>58.0</td>
<td>18.8</td>
</tr>
<tr>
<td>South Africa</td>
<td>4.9</td>
<td>9.8</td>
<td>49.6</td>
<td>22.3</td>
</tr>
<tr>
<td>Ukraine</td>
<td>8.4</td>
<td>14.6</td>
<td>57.8</td>
<td>27.8</td>
</tr>
<tr>
<td>Average</td>
<td>5.9</td>
<td>11.1</td>
<td>53.3</td>
<td>24.0</td>
</tr>
</tbody>
</table>

Source: Data tabulated by the author from Bromet et al. (2011).

High-income and low-to-middle-income countries (5.5% cf. 5.9%, respectively). The ratio of 12-month to lifetime prevalence can be taken as an indication of persistence, and therefore suggests less persistence in high-income countries. However, this result could also be due to reduced recall in low- and middle-income countries. The findings are probably not due to differential international usage of the standardized instruments, and are therefore likely to be substantive. However, interpretation in relation to local contexts is difficult.

The age of onset was really quite similar in high- and in low- to middle-income countries. However, there was an interesting and probably substantive finding in relation to age. Generally, people over 65 had the lowest rates of major depressive episode in high-income countries, but not in low- to middle-income countries, where the rates are uniform across the age groups. Kessler, Birnbaum, et al. (2010b) have shown convincingly that the decline in major depressive disorder with age cannot be due to the misattribution of affective symptoms to concomitant physical disorder. However, differential survivor bias is likely to effect comparisons between high-income countries and low- to middle-income countries. The relationship of age with prevalence is discussed further below.