

The Global Burden of Anxiety and Mood Disorders: Putting the European Study of the Epidemiology of Mental Disorders (ESEMeD) Findings Into Perspective

Ronald C. Kessler, Ph.D.

This article compares the preliminary, descriptive European Study of The Epidemiology of Mental Disorders (ESEMeD) findings reported in this supplement with 8 broad patterns of results found in previous psychiatric epidemiologic surveys. It is a systematic review of the literature on community epidemiologic surveys of anxiety and mood disorders. It concludes that the ESEMeD findings are broadly consistent with the patterns found in previous surveys but faults the preliminary ESEMeD analyses for failing to distinguish cases by severity and to consider the effects of severity on need for treatment. The fact that the ESEMeD surveys collected much richer data than previous psychiatric epidemiologic surveys on role impairment, symptom severity, and episode duration makes it possible to develop more useful classifications of clinical severity in future analyses. Elaborations that feature such distinctions have the potential to substantially increase the relevance of the ESEMeD findings for European health care policy.

(J Clin Psychiatry 2007;68[suppl 2]:10–19)

Although community epidemiologic surveys of mental disorders have been carried out in many parts of the world since shortly after World War II,^{1–3} the absence of common standards and operational procedures for diagnostic interviews has hampered efforts to make cross-national comparisons. This situation changed in the early 1980s with the development of the Diagnostic Inter-

view Schedule (DIS),⁴ the first fully structured psychiatric diagnostic interview designed for use by trained interviewers who are not clinicians. For the first time, the DIS made it possible for trained lay interviewers to carry out assessments of clinically significant mental disorders. In addition, computerized algorithms created in conjunction with the DIS made it possible to generate accurate diagnoses. The DIS was first used in the Epidemiologic Catchment Area (ECA) study,⁵ a large survey of the prevalence and correlates of mental disorders in the United States, and soon became the standard instrument for community epidemiologic surveys of mental disorders.

A number of surveys modeled on the ECA study were carried out in other countries during the 1980s.^{6–12} These surveys were subsequently brought together in a series of cross-national, comparative analyses of specific disorders.^{13–15} All of these surveys focused on *Diagnostic and Statistical Manual of Mental Disorders*, Third Edition (DSM-III) criteria, the system on which the DIS was based. However, the World Health Organization (WHO) *International Classification of Diseases (ICD)* system was used by the health care systems in a number of these countries, limiting the usefulness of the DIS findings. The WHO, recognizing this problem, developed a second, fully structured, research diagnostic interview, the WHO Composite International Diagnostic Interview (CIDI),¹⁶ which expanded the DIS to include the questions needed to make diagnoses according to the definitions and criteria of the ICD-10 system.¹⁶ In addition, the WHO coordinated the translation of the CIDI into many different languages and carried out extensive cross-national CIDI field trials to guarantee that the CIDI trans-

From the Department of Health Care Policy, Harvard Medical School, Boston, Mass.

This article is derived from the proceedings of the meeting “ESEMeD: Benefit to the Community in Depression and Anxiety,” held on September 5, 2003, in Cliveden, Berkshire, U.K. The meeting was supported by an unrestricted educational grant from GlaxoSmithKline Pharmaceuticals.

Preparation of this article was carried out in conjunction with the author’s work as the codirector, with Bedirhan Ustun, of the World Health Organization’s World Mental Health (WMH) Survey Initiative. The author thanks the staff of the WMH Data Collection and Data Analysis Coordination Centers for assistance with instrumentation, fieldwork, and consultation on data analysis. These activities were supported by the National Institute of Mental Health (R01 MH070884), the John D. and Catherine T. MacArthur Foundation, the Pfizer Foundation, the U.S. Public Health Service (R13-MH066849, R01-MH069864, and R01 DA016558), the Fogarty International Center (FIRCA R01-TW006481), the Pan-American Health Organization, Eli Lilly and Co., Ortho-McNeil Pharmaceutical, Inc., GlaxoSmithKline, and Bristol-Myers Squibb. A complete list of WMH publications can be found at <http://www.hcp.med.harvard.edu/wmh/>.

Dr. Kessler has served as a consultant to and has received grant/research support from GlaxoSmithKline.

The author thanks Kathleen Merikangas, Ph.D., National Institute of Mental Health, Bethesda, Md., and Bedirhan Ustun, M.D., World Health Organization, Geneva, Switzerland, for commenting on an earlier version of this article.

Corresponding author and reprints: Ronald C. Kessler, Ph.D., Department of Health Care Policy, Harvard Medical School, 180 Longwood Ave., Boston, MA 02115 (e-mail: kessler@hcp.med.harvard.edu).

lations yielded results of comparable reliability and validity across countries.¹⁷

Version 1.0 of the CIDI was released in 1990.¹⁸ Diagnoses were made on the basis of both the DSM-III-R and the ICD-10 criteria. The CIDI was subsequently revised to include DSM-IV criteria.¹⁹ In the decade since it first became available, the CIDI has been used in a number of large-scale, community epidemiologic surveys throughout the world.^{20–26} In recognition of this wide use, the WHO created a research consortium—the WHO International Consortium in Psychiatric Epidemiology (ICPE)—to coordinate comparative analyses. The ICPE has subsequently generated a large body of comparative cross-national data on the epidemiology of mental disorders.^{27–31}

The European Study of the Epidemiology of Mental Disorders (ESEMED) surveys are part of the generation of surveys using an expanded version of the CIDI that was built on the ICPE experience. The motivation for this CIDI expansion was the realization that the cross-national comparative analyses in the ICPE surveys were severely limited in that the original version of the CIDI consisted exclusively of diagnostic questions and basic demographic questions. No questions were included about risk factors, consequences, treatment, or barriers to seeking treatment. Although some of the ICPE surveys collected information about these other factors, the assessments were not comparable, making it impossible to carry out systematic cross-national comparisons. As additional investigators began to contact the ICPE about carrying out new CIDI surveys, a decision was made to expand the CIDI to resolve this problem by developing sections on correlates and treatment.

In addition to including entirely new sections on risk factors, consequences, treatment, and barriers to treatment, this new version of the CIDI included questions in the diagnostic sections to update the DSM assessment from version III-R to version IV. It introduced a number of impulse-control disorders to the diagnostic assessment (e.g., intermittent explosive disorder, pathological gambling, bulimia, borderline personality disorder) and retrospective assessments of several important childhood disorders (attention-deficit/hyperactivity disorder, conduct disorder, oppositional defiant disorder, and separation anxiety disorder). Furthermore, important changes in the diagnostic sections of the CIDI improved completeness and accuracy of reporting. These changes were based on the cognitive interviewing methodology strategies used in the United States National Comorbidity Survey (NCS)^{32,33} and were guided by the empirical findings in the clinical calibrations of the NCS³⁴ and the Munich version of the CIDI.³⁵ These changes improved the clinical validity of the CIDI substantially, although they reduced the comparability of this new version with the DIS and the original CIDI. A more detailed discussion of these changes is presented elsewhere.³⁶

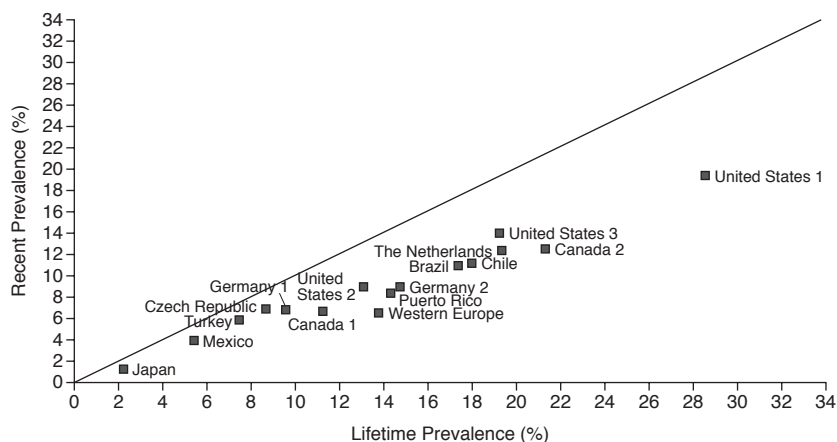
The new version of the CIDI was completed in 1999, at which time WHO established a World Mental Health (WMH) Survey Initiative, aimed at carrying out surveys with the new CIDI (WMH-CIDI) in a representative set of countries in all major regions of the world.³⁷ WMH collaborations were established in 28 different countries. The 6 ESEMED surveys are the WMH collaborating surveys in Western Europe. Additional European surveys that joined the WMH collaboration after ESEMED have either been completed or are currently being carried out in Bulgaria, Israel, Northern Ireland, Romania, Scotland, Turkey, and Ukraine. Results from these surveys will soon be available to provide additional comparisons with the ESEMED results. In addition, clinical calibration studies are currently underway to validate the WMH-CIDI in a number of countries. Three ESEMED countries are participating in this calibration exercise (France, Italy, and Spain) along with countries in all other major regions of the world (including China, Colombia, India, Mexico, Nigeria, Panama, South Africa, and the United States).

COMPARISON OF THE EUROPEAN STUDY OF THE EPIDEMIOLOGY OF MENTAL DISORDERS RESULTS WITH PREVIOUS FINDINGS

Comparison of the preliminary ESEMED results with the findings from the larger set of WMH surveys will be the most valuable comparisons in the future. However, only preliminary results from a small number of the other WMH surveys are currently available. As a result, it is necessary to turn to the published results from the early DIS surveys and the more recent ICPE surveys for comparative data. Eight broad patterns of results can be detected in the basic descriptive epidemiologic findings from these earlier surveys:

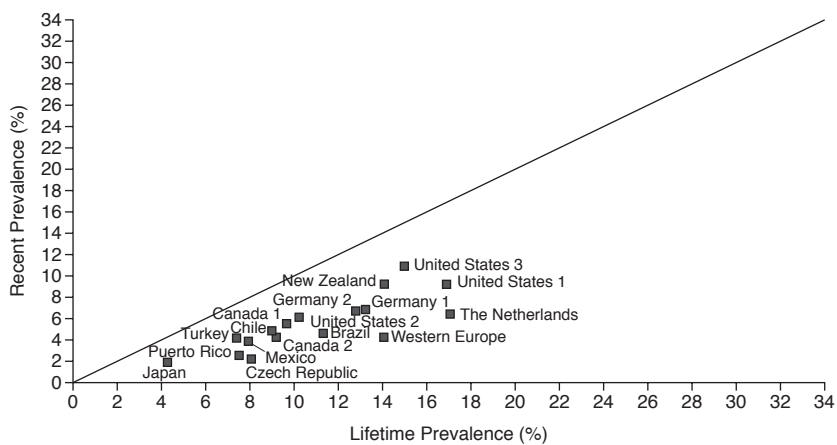
1. Wide variation has been found in these previous surveys in both lifetime and recent prevalence estimates of anxiety disorders (Figure 1) and mood disorders (Figure 2). The ESEMED lifetime prevalence estimates were in the middle of this range and recent (6- to 12-month) prevalence was somewhat lower than average.⁴¹
2. An examination of the implicit slopes in Figures 1 and 2 shows that anxiety and mood disorders are highly persistent consistently, as indicated indirectly by the ratio of 6-month or 12-month to lifetime prevalence. Roughly 60% to 70% of survey respondents with a lifetime anxiety disorder in these surveys report that their anxiety has been active in the 6 to 12 months before the interview. The comparable percentages for mood disorders are 40% to 50%. The same general pattern holds in the ESEMED data, with the persistence of anxiety disorders somewhat higher than that of mood disorders, although the

Figure 1. Lifetime and Recent (6-month or 12-month) Prevalence of Anxiety Disorders in Previous Surveys^a



^aData from: Brazil, Andrade et al.²⁶; Canada 1, Bland et al.⁶; Canada 2, Offord et al.³⁸; Chile, Vicente et al.³⁹; Czech Republic, Dragomirecka et al.⁴⁰; Germany 1, Wittchen et al.³⁵; Germany 2, Wittchen et al.¹²; Japan, Andrade et al.²⁸; Mexico, Caraveo et al.²⁴; The Netherlands, Bijl et al.²⁵; Puerto Rico, Canino et al.⁷; Turkey, Kılıç²²; United States 1, Kessler et al.²³; United States 2, Vega et al.²¹; United States 3, Robins and Regier⁵; Western Europe, Alonso et al.⁴¹

Figure 2. Lifetime and Recent (6-month or 12-month) Prevalence of Mood Disorders in Previous Surveys^a



^aData from: Brazil, Andrade et al.²⁶; Canada 1, Bland et al.⁶; Canada 2, Offord et al.³⁸; Chile, Vicente et al.³⁹; Czech Republic, Dragomirecka et al.⁴⁰; Germany 1, Wittchen et al.³⁵; Germany 2, Wittchen et al.¹²; Japan, Andrade et al.²⁸; Mexico, Caraveo et al.²⁴; The Netherlands, Bijl et al.²⁵; New Zealand, Oakley-Brown et al.⁴²; Puerto Rico, Canino et al.⁷; Turkey, Kılıç²²; United States 1, Kessler et al.²³; United States 2, Vega et al.²¹; United States 3, Robins and Regier⁵; Western Europe, Alonso et al.⁴²

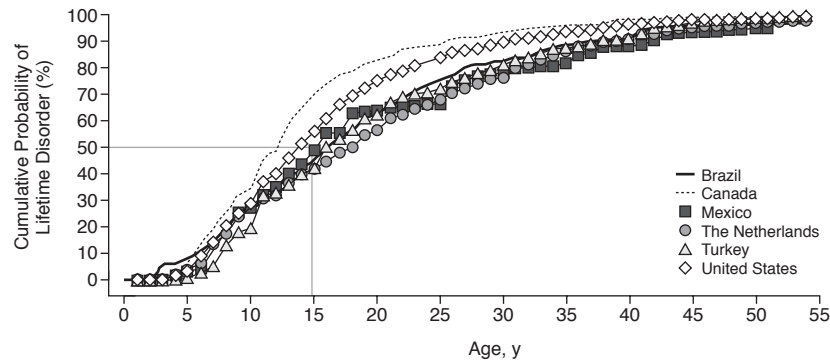
persistence ratios for both anxiety and mood disorders are lower in ESEMeD than in most of the earlier surveys.⁴¹

- Retrospective reports in earlier surveys have shown that, typically, anxiety and mood disorders have early ages at onset. Figures 3 (anxiety) and 4 (mood), which are based on the ICPE surveys, demonstrate substantial cross-national consistency in these patterns, with estimated median ages of 15 years for anxiety disorders and 26 years for mood disorders.

The ESEMeD data on age at onset have not yet been reported.

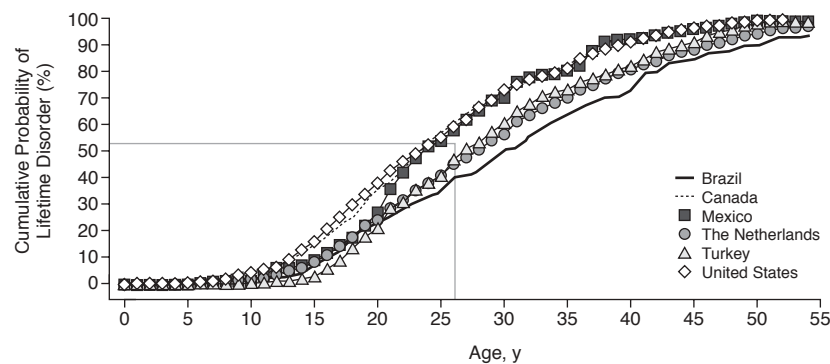
- Anxiety and mood disorders have been found in earlier surveys to be highly comorbid.^{43,44} The vast majority of people who have a history of 1 anxiety disorder typically have also been found to meet criteria for a second anxiety disorder. More than half of the people with a history of either anxiety or mood disorder have been found to have both types of disorder. Usually, anxiety has been the primary disorder

Figure 3. Age at Onset Distributions for Any Anxiety Disorders in the International Consortium in Psychiatric Epidemiology Surveys^a



^aData from Alegria et al,²⁷ Andrade et al,²⁸ Bijl et al,²⁹ WHO ICPE,³⁰ and Wang et al.³¹

Figure 4. Age at Onset Distributions for Any Mood Disorders in the International Consortium in Psychiatric Epidemiology Surveys^a



^aData from Alegria et al,²⁷ Andrade et al,²⁸ Bijl et al,²⁹ WHO ICPE,³⁰ and Wang et al.³¹

among people with a history of anxious depression. Similar patterns have been shown in epidemiologic studies of primary care samples⁴⁵ and in clinical samples of psychiatric patients.⁴⁶ The preliminary ESEMeD results have been strikingly similar to these earlier findings.⁴⁷

5. Analysis of retrospective age at onset reports in earlier surveys has shown patterns that are consistent with the suggestion that the lifetime prevalence of anxiety and mood disorders has increased in recent cohorts.³⁰ This increase is more pronounced for mood than for anxiety disorders and, among people with mood disorders, more pronounced for those with comorbid anxious depression than for those with pure depression. No data have yet been reported on whether similar patterns exist in the ESEMeD surveys.
6. Anxiety and mood disorders have been found in earlier surveys to be associated with substantial impairments in both productive roles (e.g., work absenteeism, work performance, unemployment, un-

deremployment) and social roles (e.g., social isolation, interpersonal tensions, marital disruption).^{48,49} Other evidence on this issue has come into existence over the past decade based on primary care surveys, employer disability claims data, and clinical trials. These studies, which have been reviewed elsewhere for anxiety disorders^{50,51} and mood disorders,^{31,52} document consistent, extensive role impairments associated with anxiety and mood disorders. The preliminary ESEMeD data reported in this supplement support these results.

7. In addition to being associated with the indicators of disadvantaged achievement status that follow from impairments in productive roles (e.g., low income and education), anxiety and mood disorders have also been associated with disadvantaged ascribed social status (e.g., racial-ethnic minority status, female gender). Indeed, these patterns of association have been found in epidemiologic surveys of psychological distress and mental disorders since the early 1950s.^{53,54} The ESEMeD data are consistent with

these previous results with regard to gender but are weaker than previous results regarding socioeconomic status. No ESEMeD data have yet been reported on minority status.

8. There is evidence in the ICPE surveys that delays in seeking professional treatment are widespread after first onset of an anxiety or mood disorder, especially among early-onset cases, and that only a minority of people with prevalent disorders receive any formal treatment.³⁰ We are aware of only 1 study on this issue other than in the ICPE surveys. A survey carried out by ICPE investigators among members of the Global Alliance of Mental Illness Advocacy Network asked respondents about age at disorder onset in relation to age when first receiving treatment.⁵⁵ The results were consistent with those in the ICPE surveys in finding both pervasive delays in initial help-seeking and strong positive associations between speed of initial treatment contact and age at onset. The ESEMeD surveys collected data that will allow the same patterns to be studied, but these results have not yet been reported.

DIAGNOSTIC VALIDITY

Before turning to a discussion of proposed future directions of ESEMeD analyses, it is important to comment on the validity of the ESEMeD assessment of mental disorders. There are 2 key issues in comparing results of the ESEMeD surveys with those of previous surveys. The first is the issue of diagnostic validity: whether the instruments used to operationalize DSM or ICD diagnoses in these surveys are valid. The second is the issue of practical validity: whether these diagnoses (even if valid in a narrow, technical sense of operationalizing the intended DSM or ICD criteria) are valid more broadly in identifying the range of people in need of treatment for mental disorders.

Diagnostic validity has been the focus of methodological research since the early 1980s.⁵⁶⁻⁵⁸ As of the early 1990s, Wittchen reviewed the literature in this area,¹⁷ while a number of related reports have been published subsequently that focus largely on the reliability or validity of revised versions of the CIDI.^{34,35,59-61} Three results emerge consistently from this literature:

1. The concordance between diagnoses based on the DIS or CIDI, compared with diagnoses based on blind clinical reinterviews, is far from perfect, with concordance for most diagnoses in the adequate-to-good range using conventional standards to define these characterizations.⁶² This lack of concordance is often due to 1 or 2 criteria for a particular diagnosis that are inaccurately assessed in the DIS or CIDI,⁵⁹ resulting in substantially improved concordance if these criteria are suppressed.

2. Substantial downward bias exists in the DIS and early versions of the CIDI. For example, a comparison of results collected at baseline and 1-year follow-up of the ECA study showed that a substantial number of early-onset lifetime disorders reported in the follow-up were not reported at baseline.⁶³ If we accept the well-known finding that embarrassing behaviors are much more likely to be underreported than overreported, this result is interpreted most plausibly as evidence of underreporting in the baseline interview. This bias was substantial, with up to 40% of true lifetime cases in the combined 2-wave study missed in the baseline interview. At least 2 plausible causes exist for this underreporting: the tendency for people to find it easier to recall past experiences when they are in a mood similar to how they felt in the situation they are being asked to recall; and the tendency for respondents to vary their efforts at active memory search for recall questions depending on their engagement with the interviewer and the interview. Extensive, experimental literature shows that these 2 processes have substantial effects on the accuracy of responses to recall questions.^{64,65}

3. These underreporting biases can be overcome at least partially by the use of certain strategies developed by survey methodologists to increase motivation for active memory search and to facilitate it when motivation exists.^{32,34,66} Six such strategies were used in developing the WMH-CIDI:

- A. The CIDI diagnostic stem questions for all disorders were moved to the beginning of the interview in a separate lifetime review section rather than appearing at the start of each separate diagnostic section. Such consolidation allowed memory motivation and facilitation strategies to be focused on these critical entry questions to each diagnostic section and to be administered at a point in the interview when respondents were still fresh cognitively. It also allowed all stem questions to be administered before respondents became aware that endorsement of a stem question would result in the administration of many follow-up questions.
- B. An explanation was included at the beginning of the WMH-CIDI lifetime review section aimed at increasing respondent understanding that serious memory search was required to answer the lifetime stem questions.
- C. Motivational components were included in this introduction to encourage active memory search. Specifically, respondents were told, "It is very important for the research to get complete and accurate answers to this next set of questions, so please take your time and think carefully before

answering.” Respondents were then administered a commitment probe that asked them, with this injunction as a background, if they were ready to begin. Methodological research has shown that commitment probes of this sort, in which respondents are required to acknowledge their understanding that active memory search is needed, along with their willingness to engage in it, significantly improve the accuracy of responses to survey questions that require recall.⁶⁷

- D. Interviewers were trained to read the stem questions slowly and deliberately. The aims of this approach were to make sure respondents heard all the elements of the questions, to convey the importance of the questions, and to give respondents time to begin their memory search before the questions were finished.
- E. Interviewers were trained to use feedback probes aimed at encouraging active memory search. A nondirective, reinforcing feedback probe such as “Thanks, that’s very useful,” for example, was periodically used when the respondent appeared to be taking time to think before answering. This sort of probe was used regardless of whether the respondent answered yes or no to the question. A corrective probe such as, “You answered that one awfully quickly. Are you sure there’s not something you forgot?” was used if the respondent appeared to be giving a superficial answer.
- F. The stem questions were presented to respondents on a card, as a visual aid aimed at improving question comprehension and focus.

A field experiment carried out in conjunction with the NCS showed that this set of 6 strategies, when combined, leads to a dramatic decrease in diagnostic underreporting bias as well as to an associated increase in concordance with diagnoses based on clinical reappraisal interviews.³⁴ Indeed, there were no statistically significant differences in this experiment between prevalence estimates based on the version of the CIDI that used these methodological strategies and prevalence estimates based on blind clinical interviews for 12 of 14 DSM-III-R diagnoses. Furthermore, an independent investigation showed that a single interview that adhered closely to these methodological strategies reproduced prevalence estimates obtained by combining data over 2 waves of a similar, fully structured interview that did not use these strategies.⁶³

Based on these results, one would expect, all else being equal, that the ESEMeD prevalence estimates would be more strongly concordant with independent clinical diagnoses. An ESEMeD clinical reappraisal substudy will eventually allow us to make a definitive evaluation of this expectation. One would also expect that the prevalence estimates would be higher than in previous surveys, all else

being equal. This expectation can be evaluated now. Based on the results in Figures 1 and 2, prevalence estimates in the ESEMeD surveys are roughly comparable with those of earlier surveys. In considering this result, we must recognize that the ESEMeD prevalence estimates differ in at least 1 important respect from those obtained in earlier surveys: the ESEMeD surveys, as a component of the larger WHO WMH Survey Initiative, are among the first large-scale, international, community epidemiologic surveys to base prevalence estimates on DSM-IV criteria. Previous DIS and CIDI surveys used DSM-II and DSM-III-R criteria primarily. This distinction is important because DSM-IV criteria are more strict than those in previous versions of the DSM, as a result of the more prominent emphasis on the requirement that a syndrome be associated with clinically significant distress or impairment to qualify as a disorder. Controversy exists as to the wisdom of this requirement, based on the observation that many serious medical conditions, such as hypertension and hypercholesterolemia, do not cause meaningful impairment until many years after they begin.⁶⁸ In any event, this difference means that we would expect a downward drift in ESEMeD prevalence estimates compared with earlier surveys.

Another methodological issue that may have played an important part in creating an underestimation of prevalence in the ESEMeD surveys involves fundamental survey conditions, 4 of which are especially relevant here:

1. The ESEMeD interviews differed from the typical market research interviews that the professional interviewers were used to carrying out. The ESEMeD interviews were, on the whole, longer, with an average administration time of about 2 hours. However, there was a greater range in the length of interviews than in a typical market research scenario, with the shortest taking just 45 minutes (for a respondent who denied all diagnostic stem questions) and the longest taking 4 hours (for a respondent who endorsed all stem questions and had a complex psychiatric history).
2. The interviewers were paid per interview rather than per hour and generally did not receive any additional compensation for a long interview.
3. The stem-branch structure of the CIDI creates an opportunity for interviewers to guarantee that the interview will be short merely by entering negative responses to the small number of diagnostic stem questions that guide the interview skip logic. Bias of this sort can occur either by interviewers consciously entering negative responses even when respondents answer the stem questions affirmatively or by more subtle methods that involve using voice tone, speed of reading the questions, or incorrect use of the feedback probes to induce negative responses.

4. The interviewer quality-control procedures used in the ESEMeD surveys did not guard against this type of downward-recording bias adequately. The required controls include supervisors' monitoring of the clock in the computerized software used to administer the ESEMeD interviews to make sure that diagnostic stem questions were not rushed. They also include supervisor telephone recontact of a high proportion of respondents shortly after the completion of their interviews to repeat diagnostic stem questions and make sure that responses were elicited correctly and recorded accurately. It should be noted that the ESEMeD investigators specifically contracted survey firms that had carried out surveys with similar quality-control requirements, but the firms did not implement these procedures with the rigor required to prevent bias.

Preliminary data files only became available to the ESEMeD coordinating center in Spain midway through the completion of data collection. Staff of the coordinating center detected problems of the sort described immediately when they were able to inspect the data. The ESEMeD committee made truly heroic efforts to correct the most egregious of these problems, including discarding a nearly complete survey in 1 country based on evidence that an interviewer was cheating, and conducting the survey afresh with new fundamental survey conditions designed to prevent a repeat of the same problems. However, it must be acknowledged that there is no foolproof method of guarding against subtle forms of these problems when interviewers have financial disincentives to conduct long interviews. The only truly foolproof method to deal with this problem is to pay interviewers in a way that removes any incentive to decrease the duration of the interview; paying per hour or providing a financial bonus for longer interviews are the 2 most reasonable approaches. ESEMeD used the second of these methods in some countries once it became clear that downward bias was a serious problem. It is almost certainly the case, though, that some residual downward bias remains in the data, although we have no way of knowing how large this bias might be.

PRACTICAL VALIDITY

These considerations suggest that the ESEMeD prevalence estimates might be interpreted as valid lower bounds on the true prevalence of DSM-IV disorders. Even if this tentative conclusion is accepted, though, it remains important to confront the issue of practical validity: whether these estimates define the people in need of treatment for mental disorders. Ever since the ECA study 2 decades ago estimated that one third of people in the general population of the United States meet criteria for a

DSM-III mental disorder at some time in their lives, critics have argued that these estimates are of little value for policy-planning purposes because they encompass a much higher proportion of the population than could be helped with available treatment resources.⁶⁹ As reported in the last section, we now know from numerous methodological studies that the prevalence estimates in these lay interview surveys either underestimate, or are consistent with, those obtained in blind clinical interviews. However, this fact is irrelevant to the question of practical validity, as critics who focus on practical validity call into question the overinclusiveness of the DSM and ICD systems more than the surveys that attempt to operationalize the diagnoses in these systems.

In recognition of the problem that the number of people meeting criteria for mental disorders exceeds currently available treatment resources, several more restrictive definitions have been proposed that can be used to reduce the number of people qualifying for treatment. In the United States, for example, the National Institute of Mental Health's National Advisory Mental Health Council distinguished people with Severe and Persistent Mental Illness (SPMI) from other mentally ill people and suggested that SPMI be the focus of efforts to provide universal mental health insurance coverage, regardless of ability to pay.⁷⁰ Many health plans in the United States have followed suit by restricting payment of mental health treatment to people with a subset of DSM disorders that they consider to be either very serious or "biologically based."

This is a complex issue in at least 2 ways. Firstly, it is not obvious how to define severity. Multiple dimensions of functioning are affected by mental disorders, and no clear rules exist for calibrating levels across these dimensions. In addition, current distress and functioning are not the only considerations, as patients who are currently functioning quite well nonetheless might be at extremely high risk of serious future problems based on their risk profile. Secondly, severity should not be the only criterion for deciding whether a person with a mental disorder needs treatment, as it is often possible to treat mild conditions with simple short-term treatments (e.g., behavioral treatment of specific phobia). Instead, the problem of unmet need for treatment should be addressed by developing comprehensive triage rules, which allocate available resources according to evidence-based assessments of the cost-effectiveness of available treatments across the severity threshold of the disorder. Severity gradients are widely used in this way in other branches of medicine.⁷¹ In the absence of such rules, which currently do not exist, ad hoc decision-making is inevitable.

In light of these considerations, it would be very useful for policy planning purposes if epidemiologic surveys distinguished cases that vary in severity. Consider, for example, the finding in the United States WMH survey that only 21.7% of the citizens who met criteria for DSM-IV

major depressive disorder (MDD) in the 12 months before the interview received adequate 12-month health care treatment for that illness.⁷² This low rate of treatment might be seen as a source of considerable concern by policymakers in light of evidence that MDD is a seriously impairing disorder³¹ that usually responds quite well to treatment.⁷³ However, this result is seen differently when we examine the effects of severity (as assessed by a fully structured version of a standard clinical severity scale⁷⁴) on the probability of obtaining adequate treatment (Table 1).⁷² This information shows that 39.1% of people with very severe MDD received adequate treatment, while adequate treatment was successively less common at lower levels of severity. Based on even a superficial consideration of these results, one comes to the conclusion quickly that the summary measure of 21.7% is of little value.

A great strength of the ESEMeD surveys is that they, like all the WMH surveys, contain very detailed assessments of symptom severity and role impairment. Some ESEMeD data on role impairment are presented in this supplement (see Alonso and Lepine⁷⁵). However, no results have yet been reported regarding the structured screening versions of standard symptom severity measures that were embedded in the ESEMeD surveys. Included here, for example, are structured screens that are able to assign approximate scores on the Hamilton Rating Scale for Depression⁷⁶ and the Panic Disorder Severity Scale.⁷⁷

Furthermore, we await the use of information about role impairment and symptom severity to classify cases in terms of overall clinical severity. This information is of utmost importance in the examination of treatment. While the ESEMeD results show clearly that only a minority of respondents with recent anxiety or mood disorders received treatment, we do not yet know whether this is true among people with serious or severe disorders. This kind of disaggregated analysis of treatment as a function of clinical severity should be a high priority for future ESEMeD analyses.

THE IMPORTANCE OF MILD CASES

Although we have focused on the importance of distinguishing between mild and more serious cases, it is important for the ESEMeD analyses to retain information about mild cases, even though treatment guidelines (once developed) will probably define treatment of mild cases as not being cost-effective currently. It is important to keep these mild cases in analyses both to acknowledge that mental disorders, like physical disorders, vary in severity and also to remind us that the development of cost-effective treatments for mild disorders might prevent a substantial proportion of future serious disorders.

This fact is illustrated in a recently completed investigation that was carried out in the longitudinal component of the United States NCS follow-up (NCS-2).⁷⁸ This compo-

Table 1. The Proportion of Respondents With 12-Month Major Depressive Disorder Who Received Adequate Treatment as a Function of Symptom Severity (United States World Mental Health Survey, N = 514)^a

Symptom Severity ^b	%	95% CI
Mild	12.0	2.1 to 21.8
Moderate	15.5	11.2 to 19.8
Severe	24.6	17.7 to 31.5
Very severe	39.1	29.0 to 49.2
Total	21.7	18.1 to 25.2
χ^2		23.6 ^c

^aAdapted with permission from Kessler et al.⁷²

^bSymptom severity was assessed with the Quick Inventory of Depressive Symptomatology Self-Report (Rush et al.⁷⁴).

^cThe χ^2 value measures the relationship between the percentage of respondents and the 4 levels of symptom severity; $p < .05$ (2-sided test).

nent used the WMH-CIDI to reinterview a nationally representative sample of 4375 respondents who participated a decade earlier in the baseline NCS. For purposes of this analysis, overall baseline 12-month prevalence was classified in terms of clinical severity: mild, moderate, serious, and severe cases. Over half of all cases were found to be mild. The researchers then went on to examine clinical outcomes indicative of serious mental illness over the subsequent decade, such as hospitalization for an emotional problem or a suicide attempt. Summary results are presented in Table 2.⁷⁸ There is a clear gradient in risk of serious outcomes associated with baseline clinical severity. However, risk of the outcomes is elevated among baseline mild cases compared with noncases. Furthermore, because of the much larger number of baseline mild cases than either moderate or serious cases, the population-attributable risk proportions of these outcomes due to mild cases (10.8%–12.9%) are very similar in magnitude to those associated with baseline moderate (7.8%–13.7%) and serious (10.7%–12.2%) cases.

As mentioned previously, formal cost-effectiveness analysis presumably will find that currently available treatments are not cost-effective for mild cases. This is because many forms of psychiatric treatment are more effective in reducing serious symptoms than in resolving mild symptoms and because of the fact that mild cases cause only a small reduction in quality of life, meaning that the cost savings of successful treatment will be calculated as low. The results in Table 2 remind us, though, that we need to include risk of future serious outcomes in addition to current quality of life in cost-effectiveness calculations. Furthermore, as new treatments that might be more effective and less expensive become available, cost-effectiveness results can change dramatically. The WMH-CIDI was designed with the possibility of such change in mind, recognizing that the definitions of “unmet need for treatment” will change also. This recognition was an important reason for including extensive disorder-specific data on symptom severity and role impairment in the instrument. The policy relevance of ESEMeD will be increased substantially if these data are

Table 2. The Association (odds ratios) of Baseline (1990–1992) Severity of Illness With Serious Outcomes a Decade Later (2000–2002) (National Comorbidity Survey Follow-Up, N = 4375)^{a,b}

Baseline Severity (1990–1992)	2000–2002 Outcomes														
			Hospitalization				Suicide Attempt				Any Serious Outcome ^c				
	%	SE	%	OR	95% CI	PARP	%	OR	95% CI	PARP	%	OR	95% CI	PARP	
Severe	2.9	0.4	23.8	29.7 ^d	16.9 to 52.1	32.6	8.0	11.7 ^d	4.5 to 30.4	19.5	42.4	15.1 ^d	10.0 to 22.9	17.0	
Serious	3.0	0.2	9.7	10.1 ^d	4.8 to 21.3	12.2	5.0	6.1 ^d	3.0 to 12.5	10.7	30.8	8.8 ^d	5.7 to 13.6	11.1	
Moderate	8.8	0.6	3.0	3.0 ^d	1.7 to 5.4	7.8	2.2	2.9 ^d	1.2 to 7.4	10.5	16.4	3.8 ^d	2.7 to 5.5	13.7	
Mild	16.4	0.7	2.9	2.7 ^d	1.5 to 4.9	12.9	1.6	2.0	0.8 to 4.9	10.8	9.9	2.4 ^d	1.6 to 3.4	11.8	
Noncases	68.9	0.9	1.0	1.0	0.7	1.0	4.5	1.0	
χ^2				152.1 ^d				40.4 ^d				202.8 ^d			

^aReprinted with permission from Kessler et al.⁷²

^bEntries in the % columns are unadjusted prevalences of the outcomes in subsamples defined by baseline disorder severity. Entries in the OR and 95% CI columns are obtained by exponentiating multiple logistic regression coefficients in equations that simultaneously included dummy variables for the baseline disorder severity categories and controls for age and sex to predict the outcomes. Entries in the PARP columns are associated with the severity levels of baseline disorder. A PARP is the proportion of people with a dichotomous outcome who would no longer have that outcome if the effect of the predictor variable is assumed to be causal and was set to an OR of 1.0.

^cIncludes hospitalization for an emotional problem, work disability, suicide attempt, or serious mental illness.

^dThe χ^2 value measures the relationship between serious outcomes (hospitalization, suicide attempt, or any other serious outcome) and the 5 levels of symptom severity; $p < .05$ (2-sided test).

Abbreviation: PARP = population attributable risk proportion.

Symbol: ... = not calculated.

used to refine the characterization of the descriptive epidemiology of mental disorders in Europe.

CONCLUSIONS

The brief overview presented here shows that the preliminary ESEMeD findings are, in general, consistent with previous research. There remains great potential for the ESEMeD surveys, as part of the larger WHO WMH Survey Initiative, to go beyond previous epidemiologic findings and increase the policy relevance of results. Early work of this sort has already been completed,⁷⁹ and additional work is underway. A number of key analyses still need to be carried out, especially basic comparisons of descriptive pooled cross-national data (e.g., analysis of age at onset distributions and trends in prevalence over successive cohorts) with other survey results.

Completion of the ESEMeD clinical reappraisal analyses, to compare prevalence estimates based on the fully structured WMH-CIDI with blind clinical diagnoses obtained in reinterviews with selected ESEMeD respondents, would also be useful. Ultimately, the richness of the data captured by ESEMeD using WMH-CIDI with respect to pathways and characteristics of treatment has the potential to provide important guidance regarding public health interventions that could better meet treatment needs by influencing demand for or removing barriers to treatment.

REFERENCES

- Comstock GW, Helsing KJ. Symptoms of depression in two communities. *Psychol Med* 1976;6:551–563
- Helgason T. Epidemiology of mental disorders in Iceland. *Acta Psychiatr Scand Suppl* 1964;173:115–132
- Lin TY. A study of incidence of mental disorders in Chinese and other cultures. *Psychiatry* 1953;16:315–335
- Robins LN, Helzer JE, Croughan JL, et al. National Institute of Mental Health Diagnostic Interview Schedule: its history, characteristics, and validity. *Arch Gen Psychiatry* 1981;38:381–389

- Robins LN, Regier DA. *Psychiatric Disorders in America: The Epidemiologic Catchment Area Study*. New York, NY: Free Press; 1991
- Bland RC, Orn H, Newman SC. Lifetime prevalence of psychiatric disorders in Edmonton. *Acta Psychiatr Scand Suppl* 1988;338:24–32
- Canino GJ, Bird HR, Shrout PE, et al. The prevalence of specific psychiatric disorders in Puerto Rico. *Arch Gen Psychiatry* 1987;44:727–735
- Hwu HG, Yeh EK, Cheng LY. Prevalence of psychiatric disorders in Taiwan defined by the Chinese Diagnostic Interview Schedule. *Acta Psychiatr Scand* 1989;79:136–147
- Lee CK, Kwak YS, Yamamoto J, et al. Psychiatric epidemiology in Korea, pt 1: gender and age differences in Seoul. *J Nerv Ment Dis* 1990;178:242–246
- Lépine JP, Lellouch J, Lovell A, et al. Anxiety and depressive disorders in a French population: methodology and preliminary results. *Psychiatr Psychobiol* 1989;4:267–274
- Wells JE, Bushnell JA, Hornblow AR, et al. Christchurch Psychiatric Epidemiology Study, pt 1: methodology and lifetime prevalence for specific psychiatric disorders. *Aust N Z J Psychiatry* 1989;23:315–326
- Wittchen HU, Essau CA, von Zerssen D, et al. Lifetime and six-month prevalence of mental disorders in the Munich Follow-Up Study. *Eur Arch Psychiatry Clin Neurosci* 1992;241:247–258
- Cross-National Collaborative Group. The changing rate of major depression: cross-national comparisons. *JAMA* 1992;268:3098–3105
- Weissman MM, Bland RC, Canino GJ, et al. Cross-national epidemiology of major depression and bipolar disorder. *JAMA* 1996;276:293–299
- Weissman MM, Bland RC, Canino GJ, et al. The cross-national epidemiology of panic disorder. *Arch Gen Psychiatry* 1997;54:305–309
- Robins LN, Wing J, Wittchen HU, et al. The Composite International Diagnostic Interview: an epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. *Arch Gen Psychiatry* 1988;45:1069–1077
- Wittchen HU. Reliability and validity studies of the WHO Composite International Diagnostic Interview (CIDI): a critical review. *J Psychiatr Res* 1994; 28:57–84
- World Health Organization, ed. *Composite International Diagnostic Interview (CIDI) Version 1.0*. Geneva, Switzerland: World Health Organization; 1990
- World Health Organization, ed. *Composite International Diagnostic Interview (CIDI) Version 2.1*. Geneva, Switzerland: World Health Organization; 1997
- Wittchen HU, Perkonig A, Lachner G, et al. Early developmental stages of psychopathology study (EDSP): objectives and design. *Eur Addict Res* 1998; 4:18–27
- Vega WA, Kolody B, Aguilar-Gaxiola S, et al. Lifetime prevalence of DSM-III-R psychiatric disorders among urban and rural Mexican Americans in California. *Arch Gen Psychiatry* 1998;55:771–778
- Kılıç C. *Mental Health Profile of Turkey: Main Report*. Ankara, Turkey: Ministry of Health Publications; 1998
- Kessler RC, McGonagle KA, Zhao S, et al. Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: results from the National Comorbidity Survey. *Arch Gen Psychiatry* 1994;51:8–19
- Caraveo J, Martinez J, Rivera B. A model for epidemiological studies on mental health and psychiatric morbidity. *Salud Mental* 1998;21:48–57
- Bijl RV, van Zessen G, Ravelli A, et al. The Netherlands Mental Health Sur-

- vey and Incidence Study (NEMESIS): objectives and design. *Soc Psychiatry Psychiatr Epidemiol* 1998;33:581–586
26. Andrade L, de Lolio C, Gentil V, et al. Lifetime Prevalence of Mental Disorders in a Catchment Area in Sao Paulo, Brazil. Presented at the VII Congress of the International Federation of Psychiatric Epidemiology; Aug 29–Sept 1, 1996; Santiago, Chile
 27. Alegria M, Bijl RV, Lin E, et al. Income differences in persons seeking outpatient treatment for mental disorders: a comparison of the United States with Ontario and The Netherlands. *Arch Gen Psychiatry* 2000;57:383–391
 28. Andrade L, Caraveo-Anduaga JJ, Berglund P, et al. The epidemiology of major depressive episodes: results from the International Consortium of Psychiatric Epidemiology (ICPE) Surveys. *Int J Methods Psychiatr Res* 2003;12:3–21
 29. Bijl RV, de Graaf R, Hiripi E, et al. The prevalence of treated and untreated mental disorders in five countries. *Health Aff (Millwood)* 2003;22:122–133
 30. WHO International Consortium in Psychiatric Epidemiology. Cross-national comparisons of the prevalences and correlates of mental disorders. *Bull World Health Organ* 2000;78:413–426
 31. Wang PS, Simon G, Kessler RC. The economic burden of depression and the cost-effectiveness of treatment. *Int J Methods Psychiatr Res* 2003;12:22–33
 32. Kessler RC, Mroczek DK, Belli RF. Retrospective adult assessment of childhood psychopathology. In: Shaffer D, Lucas CP, Richters JE, eds. *Diagnostic Assessment in Child and Adolescent Psychopathology*. New York, NY: Guilford Press; 1999:256–284
 33. Kessler RC, Wittchen HU, Abelson J, et al. Methodological issues in assessing psychiatric disorder with self-reports. In: Stone AA, Turkkan JS, Bachrach CA, et al, eds. *The Science of Self-Report: Implications for Research and Practice*. Mahwah, NJ: Lawrence Erlbaum Associates; 2000:229–255
 34. Kessler RC, Wittchen HU, Abelson JH, et al. Methodological studies of the Composite International Diagnostic Interview (CIDI) in the US National Comorbidity Survey. *Int J Methods Psychiatr Res* 1998;7:33–55
 35. Wittchen HU, Lachner G, Wunderlich U, et al. Test-retest reliability of the computerized DSM-IV version of the Munich-Composite International Diagnostic Interview (M-CIDI). *Soc Psychiatry Psychiatr Epidemiol* 1998;33:568–578
 36. Kessler RC, Ustun TB. The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *Int J Methods Psychiatr Res* 2004;13:93–121
 37. Kessler RC, Ustun TB. The World Health Organization World Mental Health 2000 Initiative. *Hosp Manage Int* 2000;195–196
 38. Offord DR, Boyle M, Campbell D, et al. Mental Health in Ontario: Selected Findings from the Mental Health Supplement to the Ontario Health Survey. Toronto: Queen's Printer for Ontario; 1994
 39. Vicente B, Rioseco P, Saldivia S, et al. Chilean study on the prevalence of psychiatric disorders (DSM-III-R/CIDI) (ECPD). *Rev Med Chil* 2002;130:527–536
 40. Dragomirecka E, Baudis P, Smolova E. Psychiatric morbidity of inhabitants of the Czech Republic. *Ces a Slov Psychiat* 2002;98:72–80
 41. Alonso J, Angermeyer MC, Bernert S, et al. Prevalence of mental disorders in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. *Acta Psychiatr Scand Suppl* 2004;420:21–27
 42. Oakley-Browne MA, Joyce PR, Wells JE, et al. Christchurch Psychiatric Epidemiology Study, pt 2: six month and other period prevalences of specific psychiatric disorders. *Aust N Z J Psychiatry* 1989;23:327–340
 43. Kessler RC. The prevalence of psychiatric comorbidity. In: Wetzler S, Sanderson WC, eds. *Treatment Strategies for Patients with Psychiatric Comorbidity*. New York, NY: John Wiley & Sons; 1997:23–48
 44. Merikangas KR, Angst J, Eaton W, et al. Comorbidity and boundaries of affective disorders with anxiety disorders and substance misuse: results of an international task force. *Br J Psychiatry Suppl* 1996;30:58–67
 45. Sartorius N, Ustun TB, Lecrubier Y, et al. Depression comorbid with anxiety: results from the WHO study on psychological disorders in primary health care. *Br J Psychiatry Suppl* 1996;30:38–43
 46. Maser JD, Cloninger CR. *Comorbidity of Mood and Anxiety Disorders*. Washington, DC: American Psychiatric Press; 1990
 47. Alonso J, Angermeyer MC, Bernert S, et al. 12-month comorbidity patterns and associated factors in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. *Acta Psychiatr Scand Suppl* 2004;420:28–37
 48. Kessler RC, Frank RG. The impact of psychiatric disorders on work loss days. *Psychol Med* 1997;27:861–873
 49. Kouzis AC, Eaton WW. Emotional disability days: prevalence and predictors. *Am J Public Health* 1994;84:1304–1307
 50. DuPont RL, Rice DP, Miller LS, et al. Economic costs of anxiety disorders. *Anxiety* 1996;2:167–172
 51. Greenberg PE, Sistrisky T, Kessler RC, et al. The economic burden of anxiety disorders in the 1990s. *J Clin Psychiatry* 1999;60:427–435
 52. Simon GE, Barber C, Birnbaum HG, et al. Depression and work productivity: the comparative costs of treatment versus nontreatment. *J Occup Environ Med* 2001;43:2–9
 53. Hollingshead AB, Redlich FC. *Social Class and Mental Illness: A Community Study*. New York, NY: John Wiley & Sons; 1958
 54. Srole L, Langner TS, Michael ST, et al. *Mental Health in the Metropolis: The Midtown Manhattan Study*. New York, NY: McGraw-Hill; 1962
 55. Christiana JM, Gilman SE, Guardino M, et al. Duration between onset and time of obtaining initial treatment among people with anxiety and mood disorders: an international survey of members of mental health patient advocate groups. *Psychol Med* 2000;30:693–703
 56. Anthony JC, Folstein M, Romanoski AJ, et al. Comparison of the lay Diagnostic Interview Schedule and a standardized psychiatric diagnosis: experience in Eastern Baltimore. *Arch Gen Psychiatry* 1985;42:667–675
 57. Helzer JE, Spitznagel EL, McEvoy L. The predictive validity of lay Diagnostic Interview Schedule diagnoses in the general population: a comparison with physician examiners. *Arch Gen Psychiatry* 1987;44:1069–1077
 58. Robins LN, Helzer JE, Ratcliff KS, et al. Validity of the diagnostic interview schedule, version II: DSM-III diagnoses. *Psychol Med* 1982;12:855–870
 59. Wittchen HU, Kessler RC, Zhao S, et al. Reliability and clinical validity of UM-CIDI DSM-III-R generalized anxiety disorder. *J Psychiatr Res* 1995;29:95–110
 60. Wittchen HU, Kessler RC. *The UM-CIDI: Rationale, Options, Modifications, and Technical Equivalence to the CIDI*. Munich, Germany: Max Planck Institute of Psychiatry; 1996
 61. Wittchen HU, Zhao S, Abelson JM, et al. Reliability and procedural validity of UM-CIDI DSM-III-R phobic disorders. *Psychol Med* 1996;26:1169–1177
 62. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–174
 63. Regier DA, Kaelber CT, Rae DS, et al. Limitations of diagnostic criteria and assessment instruments for mental disorders: implications for research and policy. *Arch Gen Psychiatry* 1998;55:109–115
 64. Cannell CF, Marquis KH, Laurent A. *A Summary of Studies of Interviewing Methodology*. Vital Health Statistics Series 2. No. 69. Rockville, MD: National Center for Health Statistics; 1977
 65. McFarland C, Buehler R. The impact of negative affect on autobiographical memory: the role of self-focused attention to moods. *J Pers Soc Psychol* 1998;75:1424–1440
 66. Knauper B, Cannell CF, Schwarz N, et al. Improving the accuracy of major depression age of onset reports in the US National Comorbidity Survey. *Int J Methods Psychiatr Res* 1999;8:39–48
 67. Cannell CF, Miller PV, Oksenberg L. Research on interviewing techniques. In: Leinhardt S, ed. *Sociological Methodology* 1981. San Francisco, Calif: Jossey-Bass; 1981:389–437
 68. Wakefield JC, Spitzer RL. Lowered estimates: but of what? *Arch Gen Psychiatry* 2002;59:129–130
 69. Narrow WE, Rae DS, Robins LN, et al. Revised prevalence estimates of mental disorders in the United States: using a clinical significance criterion to reconcile 2 surveys' estimates. *Arch Gen Psychiatry* 2002;59:115–123
 70. National Advisory Mental Health Council. Health care reform for Americans with severe mental illnesses: report of the National Advisory Mental Health Council. *Am J Psychiatry* 1993;150:1447–1465
 71. Summary of the second report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel II). *JAMA* 1993;269:3015–3023
 72. Kessler RC, Berglund P, Demler O, et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA* 2003;289:3095–3105
 73. Esposito K, Goodnick P. Predictors of response in depression. *Psychiatr Clin North Am* 2003;26:353–365
 74. Rush AJ, Trivedi MH, Ibrahim HM, et al. The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biol Psychiatry* 2003;54:573–583
 75. Alonso J, Lépine J-P. Overview of key data from the European Study of the Epidemiology of Mental Disorders (ESEMeD). *J Clin Psychiatry* 2007;68 (suppl 2):3–9
 76. Hamilton M. Development of a rating scale for primary depressive illness. *Br J Soc Clin Psychol* 1967;6:278–296
 77. Houck PR, Spiegel DA, Shear MK, et al. Reliability of the self-report version of the Panic Disorder Severity Scale. *Depress Anxiety* 2002;15:183–185
 78. Kessler RC, Merikangas KR, Berglund PA, et al. Mild disorders should not be eliminated from the DSM-V. *Arch Gen Psychiatry* 2003;60:1117–1122
 79. Demyttenaere K, Bruffaerts R, Posada-Villa J, et al. Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. *JAMA* 2004;291:2581–2590