Editorials

The World Health Organization
World Mental Health Survey Initiative

RONALD C. KESSLER, JOSEP MARIA HARO, STEVEN G. HEERINGA,
BETH-ELLEN PENNELL, T. BEDIRHAN ÜSTÜN

Abstract. To present an overview of the World Health Organization World Mental Health (WMH) Survey Initiative. The discussion draws on knowledge gleaned from the authors' participation as principals in WMH. WMH has carried out community epidemiological surveys in more than two dozen countries with more than 200,000 completed interviews. Additional surveys are in progress. Clinical reappraisal studies embedded in WMH surveys have been used to develop imputation rules to adjust prevalence estimates for within- and between-country variation in accuracy. WMH interviews include detailed information about sub-threshold manifestations to address the problem of rigid categorical diagnoses not applying equally to all countries. Investigations are now underway of targeted substantive issues. Despite inevitable limitations imposed by existing diagnostic systems and variable expertise in participating countries, WMH has produced an unprecedented amount of high-quality data on the general population cross-national epidemiology of mental disorders. WMH collaborators are in thoughtful and subtle investigations of cross-national variation in validity of diagnostic assessments and a wide range of important substantive topics. Recognizing that WMH is not definitive, finally, insights from this round of surveys are being used to carry out methodological studies aimed at improving the quality of future investigations.

The World Health Organization (WHO) World Mental Health (WMH) Survey Initiative is a project of the

Assessment, Classification, and Evaluation (ACE) Group at the WHO. The first phase of the project includes the implementation, coordination, and centralized analysis of data collected in a series of psychiatric epidemiological surveys carried out in more than two dozen countries around the world. Based on the rich cross-national collaborations developed in this first phase of the WMH, future phases of collaborative work are now being planned to include methodological studies, genetic epidemiological and clinical epidemiological studies, and community interventions. The current editorial focuses only on the first phase of WMH activity.

The initial rationale for the WMH project was based on the fact that government health policy decision-makers throughout the world continue to neglect mental disorders despite evidence from the WHO Global Burden of Disease (GBD) Study that mental disorders are among the most burdensome health problems in the world (Murray & Lopez, 1996). This neglect cannot be attributed to lack of evidence for effective treatments, as some mental disorders can be treated as effectively as chronic physical disorders (Bradley et al., 2005, Haby et al., 2006). However, the fact that GBD results are based largely on expert ratings of comparative illness impact rather than on empirical evidence has led to skepticism about the accuracy of estimates (Cohen, 2000, Sanderson & Andrews, 2001). The WHM surveys were designed to address this skepticism by carrying out rigorous general

population surveys that would estimate the prevalence of mental disorders and validate estimates of disease burden worldwide. The surveys were also designed to evaluate modifiable risk factors for purposes of targeting interventions and to study patterns of and barriers to service use.

The collaborating WMH countries and investigators are listed on the WMH web site at www.hcp.med.harvard.edu/wmh. The web site also includes a description of the WMH survey design procedures and quality assurance monitoring methods along with a link to the interview schedule used in the surveys. A complete list of WMH publications is also listed on the web site. Based on the easy availability of these materials, only a brief overview is presented as background in the first part of this editorial.

WMH surveys either have been completed or are in progress in six countries in the Americas, ten in Europe (including Italy) (de Girolamo *et al.*, 2005; in press), four in the Middle East, two in Africa, and four in Asia or the Pacific. The combined number of WMH interviews completed across these countries to date is in excess of 200,000. All WMH surveys use a core interview known as the WHO Composite International Diagnostic Interview (CIDI) (Kessler & Üstün, 2004) that generates diagnoses based on the definitions and criteria of both the ICD-10 and the DSM-IV systems. Survey design and implementation are coordinated by a worldwide WMH Data Collection Coordination Centre headquartered at the University of Michigan. This Centre trains the senior staff of all WMH surveys to implement a consistent set of interviewer training and quality control procedures. The Centre also carries out audits to check the quality of data collection and data cleaning.

A separate Data Analysis Coordination Centre headquartered at Harvard Medical School (HMS) carries out internal consistency checks on interviews as they are completed and entered into a computer readable format. The results of these checks are provided to supervisors to assist in quality control monitoring of interviewers and data entry specialists. Once surveys are completed, all WMH data are sent to HMS for cleaning, coding, and analysis. Cleaned and coded datasets are then returned to the participating countries for analysis. Within-country data analyses are carried out both centrally at HMS and by collaborators in the individual countries. Ongoing interaction occurs between the senior statisticians and analysts at HMS and the collaborators in the individual WMH countries to discuss statistical methods and interpretation of results, to share computer programs, and to troubleshoot various problems that arise in the course of data analysis. Peer consultation is also provided to review

drafts of papers prior to submission for publication. This process has resulted in over 50 publications from the WMH surveys as of the time of writing this editorial.

Focusing largely on broad results up to now, the published WMH reports show clearly that mental disorders are common in the vast majority of the countries studied (Demyttenaere *et al.*, 2004). A substantial proportion of these disorders are mild. However, even mild mental disorders are often found to be associated with meaningful role impairment. More severe cases are typically associated with a great deal of role impairment. Despite their high prevalence and impairment, mental disorders in most WMH countries are untreated (Von Korff *et al.*, 2005; Stang *et al.*, 2006).

UNIQUE ASPECTS OF THE WMH SURVEY ASSESSMENTS

It is noteworthy that many of the WMH surveys include a clinical reappraisal component in which a probability sample of survey respondents is interviewed by clinical interviewers who are blind to the results of the CIDI interviews. Our original hope was that these clinical interviews would be carried out in all WMH countries, but it proved impossible to do this. The clinical interviewers are carefully trained in the use of a gold standard semi-structured research diagnostic interview that is the basis for their clinical assessments. We consider this clinical reappraisal phase of the survey central to the overall WMH undertaking because it helps build a bridge between community epidemiological research and clinical practice.

Another part of this bridge-building activity involves the assessment of clinical severity. As epidemiological surveys make it clear that a substantial proportion of the population in many countries meets ICD or DSM criteria for a mental disorder, health policy planners find themselves increasingly interested in disorder severity rather than in the mere presence of a mental disorder. In order to advance our understanding of disorder severity, the WMH surveys have embedded fully structured versions of standard clinical severity measures into the assessments of specific disorders. For example, the *Quick Inventory of Depressive Symptoms Self-Report* (QIDS-SR) (Rush *et al.*, 2003) is used to assess the severity of 12-month major depressive episodes, the *Young Mania Rating scale* to assess the severity of 12-month manic episodes (Young *et al.*, 1978) and the *Panic Disorder Severity Scale* to assess the severity of 12-month panic disorder (Shear *et al.*, 2001).

A number of other features of the WMH survey assessments are also noteworthy. (i) All disorder-specific question series include a sub-threshold diagnostic assessment in order to avoid missing clinically significant syndromes by relying on rigid categorical distinctions that might not apply equally well in all countries. (ii) Special question series are included to examine overlap among diagnoses that are perhaps incorrectly separated in the current diagnostic system, such as generalized anxiety disorder and major depression. (iii) The surveys include much more information about role impairments and disability than previous psychiatric epidemiological surveys. Assessments of the impairment associated with a selected group of chronic physical conditions are also included in the survey for purposes of rank ordering the disability associated with mental disorders in comparison to the disability associated with physical disorders. (iv) The surveys include detailed questions about treatment and barriers to obtaining treatment. The information on treatment is sufficiently rich to assess adequacy of treatment. The information on barriers to treatment is sufficiently rich to help guide practical efforts at developing treatment outreach interventions.

QUALITY OF DIAGNOSTIC ASSESSMENT

Critics have rightly raised questions about the extent to which fully structured diagnostic interviews like the CIDI are capable of capturing the subtle information needed to make accurate diagnoses of ICD and DSM mental disorders (Brugha *et al.*, 1999). As it happens, we now have good information that the English version of the CIDI, in the hands of carefully trained and closely supervised lay interviewers, yields diagnoses that are quite similar to those based on independent semi-structured research diagnostic interviews carried out by experienced clinicians (Kessler *et al.*, 2004; 2005a). However, there is no guarantee that the same is true for CIDI interviews administered in other parts of the world, where the stigma associated with mental illness might be higher than in the West and where the tradition of anonymous participation in public opinion surveys is less deeply ingrained in the culture. Nor is there any guarantee, even if willingness to report emotional problems is constant across countries, that the accuracy of the translated CIDI will be equally successful in different countries in capturing the illness schemas used to make sense of the symptoms that respondents are willing to report to interviewers.

The clinical reappraisal interview component of the WMH surveys, mentioned above, was designed to address this problem of cross-national variability in the accuracy of CIDI diagnoses. We recognize that even carefully trained, experienced, culturally-competent clinicians will differ across countries in the extent to which they can accurately detect symptom-based mental disorders due to between-country differences in willingness to admit these disorders to strangers. We also recognize that variation exists in the accuracy of the reports of a single respondent over time even when that respondent is being interviewed by an expert clinical interviewer due to within-person variation in ability and willingness to recover memories and insights about symptoms that the clinical interviewer asks about. Nonetheless, for the purposes of our work we decided that the diagnoses obtained by expert clinical interviews were the standard we would try to approximate in our work.

Although WMH methodological work comparing CIDI diagnoses with clinical diagnoses is not yet complete across all countries, two general observations are worth making. First, concordance between CIDI and clinical diagnoses varies meaningfully across countries. Why this is the case is not yet entirely clear. It is apparent, though, that stigma is involved, as concordance is considerably lower for respondents who consider mental disorders stigmatizing than among those who do not. Importantly, perceived stigma varies markedly across countries. Second, it is clear that concordance between the CIDI and clinical diagnoses is significant in all countries and, importantly, that this concordance increases substantially in all countries when we use CIDI symptom data, not merely CIDI syndrome data, to predict clinical diagnoses.

To be clear on the last point: The conventional way to assess concordance between structured and semi-structured diagnoses is to examine the 2x2 table between the two diagnoses and to calculate the Kappa coefficient as a measure of concordance (Andrews & Peters, 1998). This approach fails to use all the information in the structured interviews, though, as the latter includes not only a dichotomous (yes-no) classification of each diagnosis but also dimensional information about number of symptoms and duration of symptoms and severity that could be used to improve prediction accuracy. Our CIDI clinical reappraisal analyses consistently find that the area under the receiver operating characteristic curve (AUC), a distribution-free measure of concordance that can be used in relating a predicted probability to a dichotomy, increases substantially when we add dimensional information in the CIDI to the conventional dichotomous CIDI diagnostic classification to predict clinical diagnoses.

The AUC that describes the strength of association of the CIDI in predicting clinical diagnoses varies across diagnoses within a country and across countries for a sin-

gle diagnosis. Nonetheless, in each case the parameter estimates from the prediction equation used to link information from the CIDI to clinical diagnoses in the clinical reappraisal sub-sample, when appropriately weighted, can be used to impute a probability of the clinical diagnosis to each respondent in the survey who was administered the CIDI but not the clinical reappraisal interview. Importantly, these probabilities can be used to generate unbiased estimates of the prevalence of the clinical diagnoses in the sample. These estimates, in turn, can be used to carry out epidemiological studies of clinical diagnoses based on transformation rules for converting CIDI interview data into clinical diagnoses. A detailed discussion of this approach is presented elsewhere (Kessler *et al.*, 2004). Examples of applying this approach to substantive analyses can be found in recent publications based on WMH data (Kessler *et al.*, 2005b; 2006). Even though this approach has so far been used only in within-country analyses, it can also be used to adjust for between-country differences in concordance between the CIDI and gold standard clinical interviews for purposes of carrying out rigorous cross-national comparative analyses.

THE COST-BENEFIT OF LARGE-SCALE PSYCHIATRIC EPIDEMIOLOGICAL RESEARCH

An important issue to address in evaluating the WMH is to consider whether or not a massive undertaking of this sort is cost-effective. At least one pair of critics has argued that it is not (Weich & Araya, 2004). These critics felt that the usefulness of the WMH could be called into question based on its cross-sectional design and its use of the ICD and DSM systems to classify cases. The cross-sectional design was criticized because it forces researchers to rely on potentially biased retrospective reports to make inferences about the dynamics of illness. The use of ICD and DSM categories was criticized because the validity of these categorical systems is questionable, especially in non-Western countries.

We agree with both criticisms, but not with the conclusion that the WMH is not cost-effective because of them. The problem with this conclusion is that it is based on an inaccurate assessment of the counter-balancing advantages of WMH. The two main advantages seen by the critics were (i) that the WMH surveys would generate prevalence estimates of mental disorders that could be used by policy planners and (ii) that the WMH surveys would generate estimates of the societal burden of mental disorders. The first of these two presumed advantages was criticized on the grounds that categorical models of men-

tal disorder lack validity and the second presumed advantage was criticized on the grounds that the methods used to estimate the global burden of disease in previous WHO studies have been severely criticized (Musgrove, 2003).

These criticisms can be easily dismissed. The criticism of strict adherence to invalid categorical systems is misplaced because, as noted above, the WMH questions were designed explicitly to allow for the assessment of sub-threshold cases in an explicit effort to explore the validity of the diagnostic boundaries currently specified in the ICD and DSM systems. The criticism of using controversial methods to estimate disease burden is misplaced because the main criticism of these methods has been that they rely on imputation rather than empirical analysis. The WMH surveys are carrying out precisely the kind of empirical analysis called for by the critics of previous disease burden estimates.

The critics of WMH also overlooked a number of other advantages of the initiative. (i) The infrastructure costs of the WMH Initiative were provided to participating countries without cost due to generous grants from several foundations that built the infrastructure and shared this resource through dissemination activities. This allowed participating countries to carry out high quality large-scale mental health epidemiological needs assessment surveys at a much lower cost than if they had attempted to launch such surveys on their own. In most cases, local investigators would have been unable to replicate this infrastructure regardless of cost, which means that WMH made it possible to begin a tradition of community mental health needs assessment in these countries. (ii) The baseline surveys are being used, in countries where validity was poor, as sampling frames for methodological follow-up studies aimed at improving the questions and procedures used to assess mental disorders in future research. (iii) The cross-sectional WMH surveys are serving as baselines for a number of prospective studies of exactly the sort called for by the critics. (iv) The experiences gained in WMH are helping to train a new generation of psychiatric epidemiologists in countries that lack a strong epidemiological infrastructure. This cadre of trained researchers will be of great value to healthcare planners as evidence-based methods are introduced into health policy planning in the coming years.

WHAT'S NEXT IN THE WMH SURVEY INITIATIVE?

The WMH core papers on estimated prevalence and treatment patterns are on the verge of completion in all

countries that have completed their surveys. A first WMH book reporting the results of these analyses will be completed by the end of 2006, at which time we will turn our attention to more focused analyses that will be carried out by a series of cross-national workgroups. By far the largest of these workgroups is examining patterns and consequences of comorbidity between mental and physical disorders (Von Korff *et al.*, 2005, Stang *et al.*, 2006). We anticipate that the second WMH book, scheduled for completion in 2007, will feature the results of the studies of this workgroup. Another workgroup is studying the cross-national epidemiology of non-fatal suicidal behaviors (ideation, plans, gestures, and attempts). Another is concerned with cross-national variation in the relationship between gender and psychopathology, with a special interest in the sex role experiences of women (e.g., percent of women in the labor force, access to birth control) in explaining time-space variation in gender differences. Another workgroup is examining the comparative effects of specific mental and physical disorders on disability and other aspects of disease burden. Citations to publications from these workgroups will be posted on the WMH web site as they become available for distribution.

OVERVIEW

Despite inevitable limitations imposed by the rigidity of existing diagnostic systems and by the variability of expertise in the participating countries, the WMH Survey Initiative has produced an unprecedented amount of high-quality data on the general population epidemiology of mental disorders around the world. Rather than limit themselves to the relatively mechanical analyses assumed by critics, the WMH collaborators are now actively engaged in thoughtful and subtle investigations that address the problem of cross-national variation in concordance between structured interviews and clinical assessments as well as the problem of lack of fit between diagnostic categories and experiences of psychopathology. A wide range of substantive analyses are also underway that go well beyond the simple head-counting that characterizes so much of work in this area. Finally, recognizing that WMH is not definitive, but merely a next step in the natural evolution of psychiatric epidemiology, the WMH investigators are using insights gleaned from this round of surveys to address weaknesses of the surveys by carrying out methodological studies aimed at improving the quality of future investigations.

REFERENCES

Andrews G. & Peters L. (1998). The psychometric properties of the Composite International Diagnostic Interview. *Social Psychiatry and Psychiatric Epidemiology* 33, 80-88.

Bradley R., Greene J., Russ E., Dutra L. & Westen D. (2005). A multidimensional meta-analysis of psychotherapy for PTSD. *American Journal of Psychiatry* 162, 214-227.

Brugha T.S., Bebbington P.E. & Jenkins R. (1999). A difference that matters: comparisons of structured and semi-structured psychiatric diagnostic interviews in the general population *Psychological Medicine* 29, 1013-1020.

Cohen J. (2000). The Global Burden of Disease Study: a useful projection of future global health? *Journal of Public Health Medicine* 22, 518-524.

de Girolamo G., Polidori G., Morosini G., Scarpino V., Reda V., Serra G., Mazzi F., Visonà G., Falsirollo F. & Rossi A. (2005). Prevalenza dei disturbi mentali in Italia, fattori di rischio, stato di salute ed uso dei servizi sanitari: il progetto ESEMeD-WMH. *Epidemiologia e Psichiatria Sociale*, Suppl. 8.

de Girolamo G., Polidori G., Morosini G., Scarpino V., Reda V., Serra G., Mazzi F., Alonso J., Villagut G., Visonà G., Falsirollo F., Rossi A. & Warner R. (in press). Prevalence of common mental disorders in Italy: results from the European Study of Epidemiology of Mental Disorders (ESEMeD). *Social Psychiatry and Psychiatric Epidemiology*.

Demyttenaere K., Bruffaerts R., Posada-Villa J., Gasquet I., Kovess V., Lepine J.P., Angermeyer M.C., Bernert S., de Girolamo G., Morosini P., Polidori G., Kikkawa T., Kawakami N., Ono Y., Takeshima T., Uda H., Karam E.G., Fayyad J.A., Karam A.N., Mneimneh Z.N., Medina-Mora M.E., Borges G., Lara C., de Graaf R., Ormel J., Gureje O., Shen Y., Huang Y., Zhang M., Alonso J., Haro J.M., Vilagut G., Bromet E.J., Gluzman S., Webb C., Kessler R.C., Merikangas K.R., Anthony J.C., Von Korff M.R., Wang P.S., Brugha T.S., Aguilar-Gaxiola S., Lee S., Heeringa S., Pennell B.E., Zaslavsky A.M., Üstün T.B. & Chatterji S. (2004). Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. *Journal of the American Medical Association* 291, 2581-2590.

Haby M.M., Donnelly M., Corry J. & Vos T. (2006). Cognitive behavioural therapy for depression, panic disorder and generalized anxiety disorder: a meta-regression of factors that may predict outcome. *Australian and New Zealand Journal of Psychiatry* 40, 9-19.

Kessler R.C. & Üstün T. B. (2004). The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *International Journal of Methods in Psychiatric Research* 13, 93-121.

Kessler R.C., Abelson J., Demler O., Escobar J.I., Gibbon M., Guyer M.E., Howes M.J., Jin R., Vega W.A., Walters E.E., Wang P., Zaslavsky A. & Zheng H. (2004). Clinical calibration of DSM-IV diagnoses in the World Mental Health (WMH) version of the World Health Organization (WHO) Composite International Diagnostic Interview (WMHCIDI). *International Journal of Methods in Psychiatric Research* 13, 122-139.

Kessler R.C., Berglund P., Demler O., Jin R. & Walters E.E. (2005a). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry* 62, 593-602.

Kessler R.C., Birnbaum H., Demler O., Falloon I.R.H., Gagnon E., Guyer M., Howes M.J., Kendler K.S., Shi L., Walters E. & Wu E.Q. (2005b). The prevalence and correlates of nonaffective psychosis in

the National Comorbidity Survey Replication (NCS-R). *Biological Psychiatry* 58, 668-676.

Kessler R.C., Adler L., Barkley R.A., Biederman J., Connors K., Demler O., Greenhill L., Howes M.J., Secnik K., Spencer T., Üstün T.B., Walters E.E. & Zaslavsky A.M. (2006). The prevalence and correlates of adult ADHD in the United States: results from the National Comorbidity Survey Replication. *American Journal of Psychiatry* 163(4), 716-723

Murray C.J.L. & Lopez A.D. (1996). *The Global Burden of Disease: A Comprehensive Assessment of Mortality and Disability from Diseases, Injuries and Risk Factors in 1990 and Projected to 2020*. Harvard University Press: Cambridge, MA.

Musgrove P. (2003). Judging health systems: reflections on WHO's methods. *Lancet* 361, 1817-1820.

Rush A.J., Trivedi M.H., Ibrahim H.M., Carmody T.J., Arnow B., Klein D.N., Markowitz J.C., Ninan P.T., Kornstein S., Manber R., Thase M.E., Kocsis J.H. & Keller M.B. (2003). 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. *Biological Psychiatry* 54, 573-583.

Sanderson K. & Andrews G. (2001). Mental disorders and burden of disease: how was disability estimated and is it valid? *Australian and New Zealand Journal of Psychiatry* 35, 668-676.

Shear M.K., Rucci P., Williams J., Frank E., Grochocinski V., Vander Bilt J., Houck P. & Wang T. (2001). Reliability and validity of the Panic Disorder Severity Scale: replication and extension. *Journal of Psychiatric Research* 35, 293-296.

Stang P.E., Brandenburg N.A., Lane M.C., Merikangas K.R., Von Korff M.R. & Kessler R.C. (2006). Mental and physical comorbid conditions and days in role among persons with arthritis. *Psychosomatic Medicine* 68, 152-158.

Von Korff M., Crane P., Lane M., Miglioretti D.L., Simon G., Saunders K., Stang P., Brandenburg, N. & Kessler R. (2005). A trial of an activating intervention for chronic back pain in primary care and physical therapy settings *Pain* 113, 331-339.

Weich S. & Araya R. (2004). International and regional variation in the prevalence of common mental disorders: do we need more surveys? *British Journal of Psychiatry* 184, 289-290.

Young R.C., Biggs J.T., Ziegler V.E. & Meyer D.A. (1978). A rating scale for mania: reliability, validity and sensitivity. *British Journal of Psychiatry* 133, 429-435.