Professional Psychology: Research and Practice
2010, Vol. 41, No. 6, 457–464

0735-7028/10/$12.00 DOI: 10.1037/a0021701

Toward ICD-11: Improving the Clinical Utility of WHO's International Classification of Mental Disorders

Geoffrey M. Reed
Department of Mental Health and Substance Abuse, World Health Organization, Geneva, Switzerland

This article describes the current revision by the World Health Organization (WHO) of the International Classification of Diseases and Related Health Problems (ICD-10). ICD-10 is the basis for ICD-10-CM, which will be introduced in 2013 as the official U.S. system. U.S. psychologists will be required to use ICD-10-CM for all third-party billing and reporting, but are generally not familiar with the ICD or WHO's role in global health classification. Although the U.S. lags behind other countries on the implementation of WHO's international classification systems, psychologists and other health professionals will be affected by ICD-11, so it is important to understand its development. WHO views the current revision as an important opportunity to improve the clinical utility of the classification system for mental disorders. Serious problems with the clinical utility of both the ICD and the DSM are widely acknowledged. Clinical utility affects the daily lives of practitioners and is also a global public health issue. Most people with mental disorders worldwide receive no treatment. A diagnostic system with greater clinical utility can be a tool to improve identification and treatment, helping WHO member countries to reduce the disease burden of mental disorders. Consistent with this goal, WHO's revision process is global, multilingual, and multidisciplinary and will produce different versions of the classification for clinical use, research, and primary care. A systematic program of studies being undertaken by WHO aimed at improving clinical utility is described.

Keywords: classification, ICD, mental disorders, diagnosis, clinical utility

If asked to name the current official diagnostic system for mental and behavioral disorders in the United States, most psychologists would point to the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders-IV (DSM-IV; American Psychiatric Association, 2000). In fact, this is incorrect. The official diagnostic system in the United States is based on the International Classification of Diseases and Related Health Problems (ICD) of the World Health Organization (WHO). The ICD is the global standard in diagnostic classification for health reporting and clinical applications, for mental disorders as well as for all other medical diagnoses. The current U.S. version is the ICD-9 Clinical Modification (ICD-9-CM; U.S. National Center for Health Statistics, 2010a).

All WHO member countries, including the U.S., are required by international treaty to collect and report health statistics to the WHO using the ICD as a framework. WHO is a specialized agency of the United Nations whose mission is the attainment by all peoples of the highest possible level of health and whose constitutional responsibilities include the development and maintenance of international health classification systems (World Health Organization, 2007). Progress towards better health throughout the world demands international cooperation in establishing standards and collecting and disseminating statistical information. WHO is the only organization with the ability to secure global cooperation and international agreement on these issues and is, therefore, in a unique position to initiate and promote global health standards.

WHO member countries are permitted to make certain kinds of adaptations to the ICD for country-level implementation. The U.S. government has published adaptations of ICD as the framework for official morbidity and mortality statistics for more than 50 years. The current U.S. version was adapted from the ninth revision of the ICD (ICD-9; World Health Organization, 1979) by the National Center for Health Statistics (NCHS) and the Health Care Financing Administration (now the Center for Medicare and Medicaid Services). U.S. adaptations have largely involved including additional information in order to meet hospitals' indexing needs

Editor's Note. This article is one of six articles in this special section of *Professional Psychology: Research and Practice* on "Improving Clinical Utility in the Classification of Mental and Behavioral Disorders: Revisions to the ICD and the *DSM*."—*MCR*.

GEOFFREY M. REED received his PhD in clinical psychology from the University of California, Los Angeles. He is Senior Project Officer for the Revision of ICD-10 Mental and Behavioural Disorders, Department of Mental Health and Substance Abuse at the World Health Organization in Geneva, Switzerland. He is also Adjunct Professor in the Departments of Psychology and Applied Behavioral Science, University of Kansas, and in the Department of Psychiatry, Faculty of Medicine, Universidad Autónoma de Madrid. From 1995 to 2006, he was Assistant Executive Director for Professional Development in the APA Practice Directorate. His recent publications focus on the implementation of the WHO's International Classification of Diseases and Related Health Problems (ICD) and International Classification of Functioning, Disability and Health (ICF) and on health system and professional policy related to evidence-based practice.
CORRESPONDENCE CONCERNING THIS ARTICLE should be addressed to Dr. Geoffrey M. Reed, Department of Mental Health and Substance Abuse (MER/MSD), World Health Organization, 20, Avenue Appia, CH-1211 Geneva, Switzerland. E-mail: reedg@who.int

and to enable coding of morbidity and utilization data from inpatient and outpatient records.

When ICD-9-CM was released in 1979, it was seen as a purely administrative and statistical tool, with little direct relevance to U.S. practitioners. But subsequent federal regulations required physicians to use ICD-9-CM codes on Medicare claims, and more recently, the Health Insurance Portability and Accountability Act (HIPAA) of 1996 required the use of ICD-9 CM codes on all electronic transactions for billing and reimbursement (see U.S. Centers for Medicare and Medicaid Services, 2010). Today, any psychologist who submits claims for billing and reimbursement is almost certainly using the ICD-9-CM, whether she knows it or not.

So why are mental health professionals under the impression that the DSM is the official diagnostic system? First, most U.S. psychologists were trained to use the DSM, and the current codes and coding structure of the DSM-IV and the ICD-9-CM are quite similar. So, a psychologist might use familiar code, such as 296.3 for recurrent depression, and view it as a DSM code, while the billing system recognizes it as an ICD-9-CM code. This similarity between DSM-IV and ICD-9-CM is largely a result of the American Psychiatric Association's close work with NCHS over a long period of time to promote consistency between DSM-IV and ICD-9-CM (e.g., see Thompson & Pincus, 1989). After all, there would be little use of the DSM if it failed to meet federal reporting requirements. In many cases, the American Psychiatric Association was successful in persuading the U.S. government to make the official U.S. version more similar to DSM, at the expense of compatibility with the ICD. Influencing NCHS to make the official U.S. adaptation of the ICD as compatible as possible with the DSM continues to be a focus for the American Psychiatric Association in relationship to the current development of DSM-5 (see American Psychiatric Association, 2010). Second, many institutions allow their mental health professionals to record data in clinical encounters using the DSM but then transform the data into ICD-9-CM data before submitting bills or reporting the information.

The introduction to DSM-IV-TR (American Psychiatric Association, 2000) states that its codes and terms are "fully compatible" with the ICD-9-CM (p. xxix). However, this claim is misleading. DSM-IV-TR provides a "crosswalk" to facilitate reporting of DSM-IV categories as ICD-9-CM codes. But, in spite of their general similarity, there are significant differences in the disorders included, disorder names and definitions, and the organization of categories. The differences between the DSM and the ICD-9-CM have largely to do with characteristics of WHO's ICD-9 that were retained in ICD-9-CM rather than making ICD-9-CM for mental disorders identical to DSM.

The Change That's Coming

As noted, the current U.S. system is based on the ICD-9, which was published more than 30 years ago (World Health Organization, 1979). Yet, the title of this article refers to ICD-11. So what happened to ICD-10? ICD-10 was approved by the World Health Assembly—WHO's governing body, made up of the Health Ministers of all 193 Member Countries—in 1990 (World Health Organization, 1992a). Long after its adoption by most of the rest of the world, NCHS now plans to implement the ICD-10-CM in the U.S. as of October 1, 2013 (U.S. National Center for Health Statistics, 2010b). This delay of more than two decades appears to have been substantially due to pressure from health systems and health insurers, who were resistant to changing their information systems to accommodate a new classification of diseases. The structure and coding for the mental and behavioral disorders classification in ICD-10-CM are based on ICD-10 and are substantially different from those of the ICD-9 and the DSM-IV.

As a result, psychologists and other mental health professionals will notice a big change in implementing ICD-10-CM. The familiar categories of mental and behavioral disorders are still in the system, so this will largely be a matter of learning a new arrangement of categories and a new set of alphanumeric codes. For example, in ICD-9-CM (and in DSM-IV), the code for paranoid schizophrenia is 295.3, while in ICD-10-CM, consistent with ICD-10, it is F20.0, with the "F" at the beginning of the code indicating that it is part of the ICD chapter on mental and behavioral disorders. Because the ICD-10-CM is an official activity of a U.S. government agency, there is an explicit and public process involving notices, review periods, and public hearings in order for health systems, professional groups, and other interested parties to propose changes to the ICD-10-CM. As a part of the ICD-10-CM development process, NCHS long ago announced its intention to freeze changes to the ICD-10-CM as of October 1, 2011, to allow a 2-year window to make necessary changes in health information systems and for training of health professionals and other users. For example, it would be very important for the American Psychological Association to make training resources available to its member psychologists during this time. Proposals to change the ICD-10-CM can be considered until the date of such a freeze, and the American Psychiatric Association is virtually certain to make proposals to bring ICD-10-CM more in line with its own proposals for DSM-5. There is a critical—through brief—opportunity for organized psychology to engage with NCHS to evaluate any changes proposed to ICD-10-CM to make sure they are compatible with the professional needs and clinical realities of practicing psychologists.

WHO will be completing the preparation of ICD-11 at about the same time that NCHS will be implementing ICD-10-CM. This raises the question of why the current ICD revision should be regarded as relevant by U.S. psychologists if it is going to be another 20 years before it is implemented in the U.S.. There are several reasons. First, it is highly unlikely that there will be a similar delay in ICD-11 implementation. Implementation of ICD-11 is a responsibility of the U.S. government as a WHO member state, and officials at NCHS now occupy prominent positions within the WHO Family of International Classification Network (see World Health Organization, 2010). It would be extremely difficult to justify the U.S. continuing not to use the same system that has been adopted as the standard by the rest of the world. In addition, revisions following the approval of ICD-11 will likely be made through ongoing update mechanisms (e.g., annually) rather than through periodic major revisions. To participate in this process going forward, the U.S. will need to be working from the same platform as other countries. Through a series of annual updates over the first few years following ICD-10-CM implementation, it should be possible for the U.S. to bring ICD-10-CM in line with ICD-11, so that the new system can be adopted smoothly and gradually without requiring a sudden and major change in the classification.

The Current Status of Mental Disorders Classifications

WHO sees the current revision of the ICD as an opportunity to address several major problems with current classifications of mental disorders but, most importantly, to improve its clinical utility for practitioners. The conventional wisdom is that the descriptive, criteria-based approach to diagnosis that characterizes both ICD-10 and DSM-IV has markedly improved the identification and treatment of mental disorders throughout the world (e.g., Hyman, 2007). But even the developers of these systems acknowledge that their principal achievement has been in diagnostic reliability (see Hyman, 2010). These gains are most apparent in research settings, where explicit diagnostic criteria can be applied through lengthy, complex, and costly standardized diagnostic interviews. There is no evidence of a corresponding improvement in diagnostic reliability in everyday clinical practice (see Garb, 2005).

Moreover, four decades of research have failed to demonstrate that the prevailing nosology for mental disorders is based on valid disease entities (e.g., Beutler & Malik, 2002; Charney et al., 2002). The "bootstraps" logic of the developers of the ICD and DSM classifications of mental disorders was certainly reasonable in principle (Frances, Pincus, Widiger, Davis, & First, 1990). It was expected that the improvements in reliability made possible by a descriptive, symptom-based classification would pave the way for corresponding gains in validity based on the results of clinical, neuroscience, and genetic research. This has occurred neither as rapidly nor as extensively as scientists and practitioners had hoped (Hyman, 2007; Insel et al., 2010).

Research to date has not provided a clear, validity-based overarching structure or coherent set of organizing principles for a standard diagnostic system. Nor has it led to validation of individual diagnostic entities or clarification of their criteria. Based on currently available evidence, it is safe to say that any gain in the validity of classification for ICD-11 and DSM-5 will be modest and incremental, as may be expected from an iterative process of successive approximations. In fact, Hyman (2010)—Chair of WHO's International Advisory Group for the Revision of ICD-10 Mental and Behavioural Disorders and a member of the DSM-5 Task Force—concluded that:

> (V)alidity will not be achieved simply by refining criteria for existing disorders or by the addition of new disorders. Yet DSM-IV diagnostic criteria dominate thinking about mental disorders in clinical practice, research, treatment development, and law. As a result, the modern DSM system, intended to create a shared language, also creates epistemic blinders that impede progress toward valid diagnoses (p. 155).

Clinical Utility of Current Diagnostic Systems

While conclusive demonstrations of the validity of current diagnostic systems remain elusive, there is increasing concern that current diagnostic systems are characterized by serious problems in clinical utility, even among those whose names are closely associated with their development (e.g., Andreasen, 2007; First, 2010, this issue, pp. 465–473; Kendall & Jablensky, 2003). Several major problems are highlighted in the literature. First, a high proportion of mental disorders diagnoses are recorded as "Unspecified" (the term used in the ICD) or "Not Otherwise Specified" (the corresponding DSM term). This suggests that health professionals find the categories difficult to use or not accurately descriptive of their patients or that they do not find the finer distinctions made in the classifications to be clinically useful. Second, a high proportion of people with mental health needs meet criteria for two or more disorders (e.g., Kessler, Chiu, Demler, & Walters, 2005; Krueger, Chentsova-Dutton, Markon, Goldberg, & Ormel, 2005). Much of the time, such multiple diagnoses may be considered an artifact of the classification system, representing different aspects of the same underlying condition. This suggests that existing systems are not capturing the nature of mental disorders as they are presented in clinical settings in a succinct and efficient manner.

Third, effective treatments—both psychological and pharmacological—are generally effective across a range of mental disorders (e.g., Barlow, Allen, & Choate, 2004; Gorman & Kent, 1999). This is not to suggest that the classification should be organized by treatment response or that the existence of a specific treatment should be the litmus test for inclusion of a category. But it is clear that there are many diagnostic distinctions currently made in the classification that have no relevance for clinical practice. Conversely, other diagnostic categories are characterized by significant heterogeneity that may have important treatment implications (e.g., Brinkley, Newman, Widiger, & Lynam, 2004; Parker, Fletcher, Hyett, Hadzi-Pavlovic, Barrett, & Synnott, 2008; Tackett, Krueger, Iacono, & McGue, 2005). For some disorders, it is possible for two individuals with the same diagnosis not to share *any* clinical features. For example, the DSM-IV diagnosis of conduct disorder requires the presence of as few as 3 of 15 possible criteria, and the diagnosis of substance dependence requires the presence of only 3 of 7 symptoms (American Psychiatric Association, 2000). Finally, diagnostic categories are poor predictors of treatment needs, particularly for people with the most severe forms of mental disorders, even though specific diagnoses are often used to define service eligibility for this population (Spaulding, Sullivan, & Poland, 2003). That is, existing diagnostic systems do not support the efficient use of limited treatment resources at the clinical or country level.

One of the contributors to poor clinical utility is its extraordinary complexity of current diagnostic systems, with each revision including more categories, more subtypes, and more specifiers focused on increasingly fine distinctions (see Watson, 2003). The overspecification characteristic of current mental disorders diagnostic systems has been driven partly by the nature of the revision process: incremental, framed by the existing diagnostic system, and centered on participation by psychopathology researchers with expertise in specific diagnostic groups. Watson and Clark (2006) pointed out that clinical utility is not relevant to psychopathology researchers because they are not faced with making daily treatment decisions about individual patients. They have access to increasingly sophisticated multivariate statistical techniques that can accommodate complex hierarchical models with a vast array of predictor variables. They do not wish to give up *potential* explanatory power in the development of a more nuanced scheme that *eventually* can elucidate disorder mechanisms and pathways and address issues of comorbidity and heterogeneity. The nature of current diagnostic approaches has benefited specific types of research, particularly the testing of specific pharmacotherapies for an increasing number of recognized disorders (Mays & Horwitz,

2005), but it has stifled others, such as studies of common mechanisms or treatment effectiveness across disorder groupings (Hyman, 2010; Insel et al., 2010).

An outcome of overspecification is that health professionals are commonly expected to consider between 20 and 35 separate pieces of information in diagnosing a mental disorder (Andrews, Anderson, Slade, & Sunderland, 2008). But very few medical residents and primary care physicians can remember the key symptoms for even the most common and important mental disorders (Krupinski & Tiller, 2001; Medow, Borowsky, Hillson, Woods, & Wilt, 1999). Mental health professionals may be expected to have more detailed knowledge of mental disorder symptoms, but given the extremely low rate of treatment contacts for most people with mental disorders throughout the world (Wang et al., 2007), in most settings, it is far more important to spend available time on treatment than on detailed diagnostic interviewing. Indeed, evidence is emerging that the level of detail inherent in mental disorder definitions and criteria may not be necessary. Two studies have now demonstrated that nearly half of the existing criteria for major depression could be dropped while retaining nearly complete explanatory power (Andrews et al., 2008; Zimmerman, Chelminski, McGlinchey, & Young, 2006). It is likely that similar results could be obtained for other disorders, but a systematic program of clinical and epidemiological research is needed.

Clinical Utility and WHO's Public Health Mission

Based on the abundantly evident problems with clinical utility described above, WHO has decided that improving clinical utility will be a major orienting principle of the current ICD-10 revision. Clinical utility is important to practitioners because it affects their daily lives. Moreover, WHO views clinical utility as a global public health issue.

Mental and neurological disorders account for a higher proportion of disease burden and disability than any other category of noncommunicable disease (World Health Organization, 2008). Over 75% of the disease burden of mental disorders falls on low and middle-income countries, which face the greatest dilemmas in how to direct their resources. The large majority of people with mental disorders receive no treatment or inadequate treatment. The World Mental Health Survey found that, in developed countries, between 35.5 and 50.3 percent of individuals with serious mental disorders had not received any treatment within the past month, while this was true for between 76.3 and 85.4 percent of such cases in developing countries (World Health Organization World Mental Health Survey Consortium, 2004). The "treatment gap" between those who need treatment and those who receive it is between 32% and 78%, depending on the disorder (Kohn, Saxena, Levav, & Saraceno, 2004).

In both developed and developing countries, primary care settings represent the best opportunity to improve the identification and effective treatment of people with mental disorders (e.g., Goldberg & Huxley, 1992; Wang et al., 2007). Globally, only a small minority of individuals with mental disorders—less than 10%—will ever see a psychiatrist (or, for that matter, a psychologist), and this percentage is inversely related to country income (World Health Organization, 2005). Psychiatrists are in relatively generous supply in high-income countries: about 10.5 per 100,000 population. But the proportion of the world's population that lives in these countries is small—about 15%—and declining. By contrast, there is less than one psychiatrist per 100,000 people in low-income countries, and an average of only about two psychiatrists per 100,000 in middle-income countries. To serve public health needs, the diagnostic system must be usable in primary care settings by health care providers who are not mental health specialists and who, in lower-resource countries, may have limited formal professional training of any kind.

People are only likely to have access to the most appropriate mental health services when the conditions that define eligibility and treatment selection have their basis in a diagnostic system that is usable and useful in the places where people with mental disorders are most likely to come into contact with the health care delivery system. Data generated in the context of health encounters are used by governments for many purposes, including morbidity reporting to WHO, allocation of health care resources, the development of clinical and public health programs, and as a basis for health policy, such as in eligibility determination and reimbursement. If a diagnostic system is characterized by poor clinical utility at the encounter level, it cannot generate data based on those encounters that will be a valid basis for health programs and policies. That is, for the validity of health encounter data, clinical utility is as much a requirement as reliability.

The public health implications of clinical utility have shaped WHO's approach to the ICD revision in several important ways. First, to impact public health, the development of ICD-11 must be global. Like the U.S., other countries have developed modifications of ICD to reflect their information needs and priorities for health. An effective multilateral process must give users in other countries—through legitimate representatives and jointly determined mechanisms—a meaningful opportunity to shape the final product. Securing strong participation from developing countries is a challenge and a requirement. A diagnostic system with high clinical utility for these countries that can be a tool for helping to reduce disease burden cannot be created simply by "dumbing down" a research classification after it has been developed.

A second and closely related issue is that the revision process and the revision products must be multilingual. Many of the specificities of cultural and country perspectives are embedded in language, and if there is not attention to translation and linguistic equivalence until the end of the process, reduced clinical utility for non-English versions is a predictable result. Specific WHO activities related to developing the ICD-11 classification of mental and behavioral disorders are being conducted in more than 20 languages. WHO will publish core ICD-11 mental and behavioral disorders classification products in all WHO official languages: Arabic, Chinese, English, French, Russian, and Spanish.

Third, the revision effort must be multidisciplinary. As noted, the vast majority of people with mental disorders worldwide will never see a psychiatrist. In order to serve as a tool to reduce disease burden, the ICD-11 classification of mental and behavioral disorders will need to be usable by a much broader range of health care personnel. WHO views all health professionals who use the mental disorders classification as a constituency for its revision. This is reflected in the prominent representation in WHO's revision process of several international professional societies with a legitimate claim to global representation in their respective disciplines (e.g., psychology, social work, nursing, primary care medicine).

Finally, WHO believes that one size does *not* fit all. WHO published three versions of the ICD-10 classification of mental and behavioral disorders: 1) the *Clinical Descriptions and Diagnostic Guidelines* (World Health Organization, 1992b), primarily intended for mental health professionals; 2) the *Research Diagnostic Criteria* (World Health Organization, 1993); and 3) *Diagnostic and Management Guidelines for Mental Disorders in Primary Care* (World Health Organization, 1996). The *Clinical Descriptions and Diagnostic Guidelines* provides clinician-oriented guidance that includes narrative paragraphs that are similar to prototypes and diagnostic guidance that is sufficiently flexible to allow for cultural variation and clinical judgment. The *Research Diagnostic Criteria* are fully operationalized for use in clinical and epidemiological studies (see First, 2009, for a comparison with DSM-IV). The primary care classification is a simplified system consisting of 27 categories that relate to common presentations in general medical care settings in a format intended to be accessible to professionals in those settings and also includes basic management guidance. WHO's experience and reports from the field are generally that these different products are perceived as valuable and suited to their different purposes. WHO intends to produce analogous products for ICD-11.

Operationalizing Clinical Utility

Clinical utility is mentioned often in discussions of the goals and purposes of classification systems for mental disorders. For example, DSM-IV-TR explicitly states that its highest priority is "to provide a helpful guide to clinical practice" (American Psychiatric Association, 2000, p. xxiii), but its developers admit that the nearly exclusive focus of its development was diagnostic validity (First et al., 2004). In contrast, the ICD-10 clinical descriptions and diagnostic guidelines for mental and behavioral disorders (World Health Organization, 1992b) was tested in a clinical utility field trial involving more then 15,000 individual assessments conducted by more than 700 clinicians in 39 countries (Sartorius et al., 1993). For each diagnostic assessment, clinicians were asked to rate: a) the goodness of fit of the ICD-10 clinical descriptions and diagnostic guidelines; b) their confidence in using the diagnosis; c) the ease or difficulty of making the diagnosis using the manual; and d) the adequacy of clinical descriptions and diagnostic guidelines. Operational definitions were provided for each point on the rating scales for each dimension. Overall, clinicians reported that the draft document was easy to use and that the classification provided a good fit for most of the clinical conditions encountered.

This WHO field trial, though an important step, was a test of a draft system that had already been developed. Until quite recently, there has been very little attention to the issue of how a classification system might be constructed with the explicit goal of maximizing clinical utility—that is, what its architecture, categories, and descriptive characteristics should be in order to make it as useful as possible to clinicians.

In order to use clinical utility as an organizing principle in classification development, it is necessary to define and operationalize it. Mullins-Sweatt and Widiger (2009) provided an overview of conceptualizations of clinical utility, pointing out that developers of both the ICD and the DSM conflated utility and validity in their discussions of the issue. The most widely cited definition of clinical utility was proposed by First et al. (2004). (See First, 2010, this issue, for a revised definition.) Mullins-Sweatt and Widiger (2009) argued that specific aspects of the First et al. (2004) definition—specifically, the usefulness of a particular classification concept or category in *conceptualizing* the disorders and its value in *predicting* future management needs—actually related to construct and predictive validity, respectively, and proposed a narrower definition.

The WHO Department of Mental Health and Substance Abuse is using a working definition of clinical utility in developing the ICD-11 classification of mental and behavioral disorders that is somewhat simpler than First et al. (2004) definition but does not adhere to Mullins-Sweat and Widiger's (2009) stricter separation of utility and validity. According to our working definition, the clinical utility of a classification construct or category for mental and behavioral disorders depends on: a) its value in *communicating* (e.g., among practitioners, patients, families, administrators); b) its *implementation characteristics* in clinical practice, including its goodness of fit (i.e., accuracy of description), its ease of use, and the time required to use it (i.e., feasibility); and c) its usefulness in *selecting interventions* and in making *clinical management* decisions. We would also include in a definition of clinical utility the extent to which the use of the construct or category is associated with improvements in clinical outcomes at individual level and in health status at population level, though we do not anticipate being able to test the classification's impact on longitudinal outcomes as a part of the current ICD revision process. While we acknowledge that most of the above points can also be discussed in terms of validity, these are very different validity characteristics than have previously been considered as a part of classification development.

Using Clinical Utility Studies as a Basis for ICD-11

The WHO Department of Mental Health and Substance Abuse is currently undertaking a programmatic series of investigations regarding how the ICD-11 classification of mental and behavioral disorders diagnostic system can be constructed to maximize clinical utility. Broadly speaking, these clinical utility studies can be divided into two types: formative field studies and evaluative field studies. Formative field studies are those undertaken early in the development process in order to inform decisions about the basic structure and content of the classification. This includes information about the overall structure, content, or approach that users will find useful. Evaluative field studies are those conducted after a draft of the classification has been produced to look at various aspects of its application. This includes studies of whether improvements in clinical utility have been achieved.

To assist in the development and implementation of formative field studies, the Department of Mental Health and Substance Abuse has appointed a Formative Field Studies Coordination Group and has established a global network of Formative Field Study Centers. These include Field Study Centers in Brazil, China, India, Mexico, and Nigeria. Between them, these five low and middle-income countries represent nearly 45% of the world's population. Field Study Centers have also been established in Japan, Spain, and the U.S., and a Center for the Arab region is in the process of being established.

Based on clinical utility as an overarching priority, formative field studies will address three broad questions:

1. How Should Disorder Categories Be Organized?

Several proposals have been made for reorganization of existing diagnostic categories, based on different interpretations of available validity evidence (e.g., Andrews et al. 2009). Unfortunately, existing validity evidence does not provide uniform or definitive support for one particular architecture. Even if it did, such an organization might not be most useful clinically. For example, the knowledge that a particular genetic marker is shared across two disorders will not be clinically relevant if this information is not available for individual patients in clinical settings or if clinical management issues and treatment strategies for the two disorders are different in spite of their shared genetic substrate.

Discussions to date about the classification "metastructure" have been based on the perspective that what is most important in creating the structure of the classification is reflecting the scientific "truth" of the relationships among disorders. However, it may be unwise to expect that there would be a single and entirely satisfactory way to map a set of complex, interrelated, overlapping, largely continuous phenomena with poorly understood biological bases that interact with development and the environment in multiple ways onto a discrete set of large category groupings that will serve as the basic structure of a classification for all settings and uses. Attending only to "validity indicators" (see Hyman, 2010) would inevitably lead to a situation where the classification structure must be based on mixed, conflicting, incomplete, or absent evidence. In the end, a fairly high proportion of the resulting decisions would be arbitrary.

However, another way to think about the architecture of the classification (i.e., the Table of Contents or larger groupings of disorders) is as a sort of user interface that should allow the person using the system to find the category that he or she wishes to use—that fits the case in front of him or her—as quickly, easily, and intuitively as possible. If one thinks about the classification "metastructure" in this way, it seems possible that it may not need to be the same for all types of users and for all classification purposes. As long as the classification categories themselves are equivalent, it may not matter that a genetics researcher finds one organization of categories most useful and a community health worker another or that a different organization is more intuitive in Japan than in the U.S.. This can be viewed as an empirical question about the usability of the classification by its intended users, without getting into an argument about the relative scientific validity of alternative organizations. Of course, sometimes clinicians' opinions may simply be wrong. If there is strong evidence in support of an approach at odds with their views, then the classification should be structured accordingly and a corrective educational effort initiated. In the absence of such dispositive evidence; however, the extent to which alternative organizations facilitate the use of the classification in clinical settings should weigh strongly in decisions about the final structure.

Flanagan, Keeley, and Blashfield (2008) suggested an interesting avenue of investigation on developing a more clinically intuitive classification architecture by using a series of methodologies derived from anthropology and cognitive science to illuminate clinicians' working taxonomies of mental disorders. They found that clinicians' mental taxonomies are reliable, sensible, and meaningful and are largely based on distinctions that are important for management and treatment. (See Flanagan and Blashfield, 2010, this issue, pp. 474–481, for a more recent summary and update of this line of research.) Two field studies protocols, based on methods from psychology, cognitive science, and anthropology, including a partial replication and multilingual expansion of Flanagan et al.'s 2008 study, are currently being implemented internationally to inform decisions about the overall architecture of the system by enabling examination of the relationship of possible structures to clinicians' conceptualizations of mental disorders and their clinical management.

2. What Disorders, Conditions, or Problems Should Be Included in a Diagnostic System in Order To Facilitate Appropriate Identification and Treatment of Mental and Behavioral Disorders at Each Level of Care?

This is a particularly important issue in primary care settings, where the classification must be limited to a much smaller number of categories if is has any hope of being widely used. However, the primary care version cannot be generated simply by using the Table of Contents entries from the specialist classification (Gask, Klinkman, Fortes, & Dowrick, 2008). The most useful categories in primary care may be ones that combine elements that are in different clusters in the specialist or research versions. For example, at the specialist level and in many research projects, there may be good reasons for thinking about anxiety disorders and depression differently. In primary care, a broader concept of emotional disorders, combining these two elements, may be most useful. Similarly, there are meaningful distinctions at the specialist level and in research between conduct disorder, oppositional-defiant disorder, and psychopathy among children, but these may all be more helpfully characterized under the heading of disruptive behavior at the primary care level. Developing the primary care classification will involve specifying and testing classification entities that may be different from entities used by specialists, as well as developing algorithms for cross-linkage of primary care, specialist, and research categories.

One source of relevant information will be data on current use of classification categories to the extent that is available. It has been suggested that many of the available mental disorders diagnoses are never or almost never used (Müssigbrodt et al., 2000). However, most of the available data bearing on this issue is from inpatient hospitals, leading to substantial diagnostic bias. For this reason, new data collections in targeted settings will be implemented. Studies in primary care settings will focus on obtaining the best coverage of high-incidence and high-resource mental health problems, some of which may be most usefully described at the level of problems or symptom clusters rather than by formal diagnoses. Studies in specialty mental health settings will focus on the extent to which categories and diagnostic distinctions are relevant to clinicians in assessment, problem conceptualization, treatment selection and management, and outcomes of services. At both levels, studies will focus on whether a simpler system with fewer categories may be more useful in the identification and appropriate treatment of individuals with mental health needs. While it is often assumed that researchers will prefer the highest possible level of detail in mental and behavioral disorders classification, this is an empirical question. The perceived utility of the system for researchers could be tested in a similar manner.

3. How Should the Information for Each Disorder Be Presented?

A number of alternatives to existing diagnostic approaches have been proposed as ways to improve the clinical utility of diagnostic systems, including dimensional ratings (e.g., Widiger & Mullins-Sweatt, 2010, this issue, pp. 488–494) and prototype matching (e.g., Westen, DeFife, & Bradley, & Hilsenroth, 2010, this issue, pp. 482–487). A mix of models will likely be required to classify, diagnose, and inform management decisions related to mental and behavioral disorders. Studies will focus on the comparative clinical utility of different approaches to disorder descriptions and diagnostic guidelines in specific settings and for different purposes.

Practitioner Surveys

If a core goal of the ICD revision is to improve the classification's utility for its users, then it seems important to ask them directly about their views and experiences of existing classification systems. The WHO Department of Mental Health and Substance Abuse is collaborating with the World Psychiatric Association (WPA) and WPA member societies (psychiatric associations at the national level) to conduct an international and multilingual survey of psychiatrists' views and experiences of ICD-10. To date, 54 WPA Member Societies from 50 countries are participating. Of these, 30 are low and middle-income countries. The survey is being conducted in more than 20 languages.

The survey is conducted via the internet. Responses to date suggest minimal difficulty with online participation, even in developing countries. For example, over three-quarters of the members of the Indian Psychiatric Society have participated in the survey to date. The survey focuses on key practical and conceptual issues, such as the most important purpose of a classification, desired number of categories, desirability of a strict criteria-based approach in contrast to a more flexible one, the best way to incorporate severity and disability, and cultural applicability. Respondents are asked to provide ratings of ease of use and goodness of fit of diagnostic categories they report using at least once per week. The survey is specifically intended to assess the perspectives of "rank and file" practitioner psychiatrists, rather than those of the Society's leadership, the members of a particular section or division, or putative classification experts. Although this seems relatively straightforward, nothing of this international scope has ever been undertaken. WHO would like to conduct a similar survey of the views and experiences of other groups of health professionals, including psychologists, but this will require that psychological associations, including APA, organize themselves in order to be able to participate.

Conclusions

WHO's focus on clinical utility in the development of a revised classification of mental and behavioral disorders has the potential to produce a system that is better tied to the needs of real-life clinical practice for a variety of health professionals, including psychologists. Although many psychologists are not aware of the relevance of the ICD to their practices, the ICD-11 will have a direct effect on the official classification used in the U.S., which must conform to the ICD for the purposes of health statistics and reporting. WHO looks forward to the participation of American psychology and our international and multidisciplinary colleagues in creating a diagnostic system for mental disorders with high clinical utility that supports the delivery of much-needed services to people with mental health needs and is consistent with international scientific evidence and the mission of WHO. Such a system will help to align psychological practice with both science and the public health agenda, and provide a part of the foundation for an integrated model of psychological practice in the future.

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

American Psychiatric Association. (2010). *DSM-5 development: Timeline.* Retrieved from http://www.dsm5.org/about/Pages/Timeline.aspx

Andreasen, N. C. (2007). DSM and the death of phenomenology in America: An example of unintended consequences. *Schizophrenia Bulletin, 33,* 108–112.

Andrews, G., Anderson, T., Slade, T., & Sunderland, M. (2008). The classification of anxiety and depressive disorders: Problems and solutions. *Depression and Anxiety, 25,* 274–281.

Andrews, G., Goldberg, D. P., Krueger, R. F., Carpenter Jr., W. T., Hyman, S. E., Sachdev, P., & Pine, D. (2009). Exploring the feasibility of a meta-structure for DSM-V and ICD-11: Could it improve utility and validity? *Psychological Medicine, 39,* 1993–2000.

Barlow, D. H., Allen, L. B., & Choate, M. L. (2004). Toward a unified treatment for emotional disorders. *Behavior Therapy, 35,* 205–230.

Beutler, L. W., & Malik, M. L. (Eds.). (2002). *Rethinking the DSM: A psychological perspective.* Washington, DC: American Psychological Association.

Brinkley, C. A., Newman, J. P., Widiger, T. A., & Lynam, D. R. (2004). Two approaches to parsing the heterogeneity of psychopathy. *Clinical Psychology: Science and Practice, 11,* 69–94.

Charney, D. S., Barlow, D. H., Botteron, K., Cohen, J. D., Goldman, D., Gur, R. E., & Zalcman, S. J. (2002). Neuroscience research agenda to guide development of a pathophysiologically based classification system. In D. E. Kupfer, M. B. First, & D. A. Regier (Eds.), *A research agenda for DSM-V* (pp. 31–83). Washington, DC: American Psychiatric Association.

First, M. B. (2009). Harmonization of DSM-V and ICD-11: Opportunities and challenges. *British Journal of Psychiatry, 195,* 382–390.

First, M. B. (2010). Clinical utility in the revision of the Diagnostic and Statistical Manual of Mental Disorders (DSM). *Professional Psychology: Research and Practice, 41,* 465–473.

First, M. B., Pincus, H. A., Levine, J. B., Williams, J. B. W., Ustun, B., & Peale, R. (2004). Clinical utility as a criterion for revising psychiatric diagnoses. *American Journal of Psychiatry, 161,* 946–954.

Flanagan, E., Keeley, J., & Blashfield, R. K. (2008). An alternative hierarchical organization of the mental disorders of the DSM-IV. *Journal of Abnormal Psychology, 117,* 693–698.

Flanagan, E. H., & Blashfield, R. K. (2010). Increasing clinical utility by aligning the DSM and ICD with clinicians' conceptualizations. *Professional Psychology: Research and Practice, 41,* 474–481.

Frances, A. J., Pincus, H. A., Widiger, T. A., Davis, W. W., & First, M. B. (1990). DSM-IV: Work in progress. *American Journal of Psychiatry, 147,* 1439–1448.

Garb, H. (2005). Clinical judgment and decision making. *Annual Review of Clinical Psychology, 1,* 167–195.

Gask, L., Klinkman, M., Fortes, S., & Dowrick, C. (2008). Capturing complexity: The case for a new classification system for mental disorders in primary care. *European Psychiatry, 23,* 469–476.

Goldberg, D. P., & Huxley, P. J. (1992). *Common mental disorders: A bio-social model.* London, England: Tavistock.

Gorman, J. M., & Kent, J. M. (1999). SSRIs and SNRIs: Broad spectrum of efficacy beyond major depression. *Journal of Clinical Psychiatry, 60,* 33–39.

Hyman, S. E. (2007). Can neuroscience be integrated into the DSM-V? *Nature Reviews: Neuroscience, 8,* 725–232.

Hyman, S. E. (2010). The diagnosis of mental disorders: The problem of reification. *Annual Review of Clinical Psychology, 6,* 155–179.

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., & Wang, P. (2010). Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. *American Journal of Psychiatry, 167,* 748–751.

Kendall, R., & Jablensky, A. (2003). Distinguishing between the validity and utility of psychiatric diagnoses. *American Journal of Psychiatry, 160,* 4–12.

Kessler, R. C., Chiu, W. T., Demler, O., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry, 62,* 617–627.

Kohn, R., Saxena, S., Levav, I., & Saraceno, B. (2004). Treatment gap in mental health care. *Bulletin of the World Health Organization, 82,* 858–866.

Krueger, R. F., Chentsova-Dutton, Y. E., Markon, K. E., Goldberg, D., & Ormel, J. (2003). A cross-cultural study of the structure of comorbidity among common psychopathological syndromes in the general health care setting. *Journal of Abnormal Psychology, 107,* 216–227.

Krupinski, J., & Tiller, J. W. G. (2001). The identification and treatment of depression by general practitioners. *Australian and New Zealand Journal of Psychiatry, 35,* 827–832.

Mays, R., & Horwitz, A. V. (2005). DSM-III and the revolution in the classification of mental illness. *Journal of the History of the Behavioral Sciences, 41,* 249–267.

Medow, A., Borowsky, S. J., Dysken, S., Hillson, S. D., Woods, S., & Wilt, T. J. (1999). Internal medical resident's ability to diagnose and characterise major depression. *Western Journal of Medicine, 17,* 35–40.

Mullins-Sweatt, S. N., & Widiger, T. A. (2009). Clinical utility and DSM-V. *Psychological Assessment, 21,* 302–312.

Müssigbrodt, H., Michels, R., Malchow, C. P., Dilling, H., Munk-Jørgensen, P., & Bertelsen, A. (2000). Use of the ICD-10 classification in psychiatry: An international survey. *Psychopathology, 33,* 94–99.

Parker, G., Fletcher, K., Hyett, M., Hadzi-Pavlovic, D., Barrett, M., & Synnott, H. (2008). Measuring melancholia: The utility of a prototypic symptom approach. *Psychological Medicine, 39,* 989–998.

Sartorius, N., Kaelber, C. T., Cooper, J. E., Roper, M. T., Rae, D. S., Gulbinat, W. M., & Regier, D. A. (1993). Progress toward achieving a common language in psychiatry: Results from the field trial of the clinical guidelines accompanying the WHO classification of mental and behavioral disorders in ICD-10. *Archives of General Psychiatry, 50,* 115–224.

Spaulding, W. D., Sullivan, M. E., & Poland, J. S. (2003). *Treatment and rehabilitation of severe mental illness.* New York, NY: Guilford.

Tackett, J. L., Krueger, R. F., Iacono, W. G., & McGue, M. (2005). Symptom-based subfactors of DSM-defined conduct disorder: Evidence for etiologic distinctions. *Journal of Abnormal Psychology, 114,* 483–487.

Thompson, J. W., & Pincus, H. (1989). A crosswalk from DSM-III-R to ICD-9-CM. *American Journal of Psychiatry, 146,* 1315–1319.

U.S. Centers for Medicare and Medicaid Services. (2010). *Transaction and code sets standards.* Retrieved from http://www.cms.gov/TransactionCodeSetsStands/

U.S. National Center for Health Statistics. (2010a). *International classification of diseases* (9th rev., Clinical modification). Retrieved from http://www.cdc.gov/nchs/icd/icd9cm.htm

U.S. National Center for Health Statistics. (2010b). *International classification of diseases* (10th rev., Clinical modification). Retrieved from http://www.cdc.gov/nchs/icd/icd10cm.htm

Wang, P. S., Aguilar-Gaxiola, S., Alonso, J., Angermeyer, M. C., Borges, G., Bromet, E. J., & Wells, J. E. (2007). Worldwide use of mental health services for anxiety, mood, and substance disorders: Results from 17 countries in the WHO world mental health (WMH) surveys. *Lancet, 370,* 841–850.

Watson, D. (2003). Subtypes, specifiers, epicycles, and eccentrics: Toward a more parsimonious taxonomy of psychopathology. *Clinical Psychology: Science and Practice, 10,* 233–238.

Watson, D., & Clark, L. A. (2006). Clinical diagnosis at the crossroads. *Clinical Psychology: Science and Practice, 13,* 210–215.

Westen, D., DeFife, J. A., Bradley, B., & Hilsenroth, M. J. (2010). Prototype personality diagnosis in clinical practice: A viable alternative for DSM-V and ICD-11. *Professional Psychology: Research and Practice, 41,* 482–487.

Widiger, T. A., & Mullins-Sweatt, S. N. (2010). Clinical utility of a dimensional model of personality disorder. *Professional Psychology: Research and Practice, 41,* 488–494.

World Health Organization. (1979). *International statistical classification of diseases, injuries, and causes of death* (9th rev.). Geneva, Switzerland: Author.

World Health Organization. (1992a). *International classification of diseases and related health problems* (10th rev.). Geneva, Switzerland: Author.

World Health Organization. (1992b). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines.* Geneva, Switzerland: Author.

World Health Organization. (1993). *The ICD-10 classification of mental and behavioural disorders: Diagnostic criteria for research.* Geneva, Switzerland: Author.

World Health Organization. (1996). *Diagnostic and management guidelines for mental disorders in primary care: ICD-10 Chapter V Primary Care Version.* Göttingen, Germany: Holgrefe.

World Health Organization. (2005). *Mental health atlas 2005.* Geneva, Switzerland: Author.

World Health Organization. (2007). *Basic documents* (46th ed.). Geneva, Switzerland: Author.

World Health Organization. (2008). *The global burden of disease: 2004 update*. Geneva, Switzerland: Author.

World Health Organization. (2010). *The WHO-FIC network.* Retrieved from http://www.who.int/classifications/network/en/

World Health Organization World Mental Health Survey Consortium. (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.

Zimmerman, M., Chelminski, I., McGlinchey, J. B., & Young, D. (2006). Diagnosing Major Depressive Disorder: X. Can the utility of the DSM-IV symptom criteria be improved? *Journal of Nervous and Mental Disorders, 194,* 893–897.

Received August 3, 2010
Revision received September 20, 2010
Accepted September 20, 2010 ■