

# Developing a Science of Clinical Utility in Diagnostic Classification Systems Field Study Strategies for *ICD-11* Mental and Behavioral Disorders

Jared W. Keeley  
Mississippi State University

Geoffrey M. Reed  
World Health Organization, Geneva, Switzerland, and  
Universidad Nacional Autónoma de México

Michael C. Roberts and Spencer C. Evans  
University of Kansas

María Elena Medina-Mora and Rebeca Robles  
National Institute of Psychiatry “Ramón de la Fuente”

Tahilia Rebello  
Columbia University Medical Center and New York State  
Psychiatric Institute

Pratap Sharan  
All India Institute of Medical Sciences

Oye Gureje  
University of Ibadan

Michael B. First and Howard F. Andrews  
Columbia University Medical Center and New York State  
Psychiatric Institute

José Luís Ayuso-Mateos  
Universidad Autónoma de Madrid

Wolfgang Gaebel and Juergen Zielasek  
Heinrich-Heine University

Shekhar Saxena  
World Health Organization, Geneva, Switzerland

The World Health Organization (WHO) Department of Mental Health and Substance Abuse has developed a systematic program of field studies to evaluate and improve the clinical utility of the proposed diagnostic guidelines for mental and behavioral disorders in the Eleventh Revision of the *International Classification of Diseases and Related Health Prob-*

Jared W. Keeley, Department of Psychology, Mississippi State University; Geoffrey M. Reed, Department of Mental Health and Substance Abuse, World Health Organization, Geneva, Switzerland and Faculty of Psychology, Universidad Nacional Autónoma de México; Michael C. Roberts and Spencer C. Evans, Clinical Child Psychology Program, University of Kansas; María Elena Medina-Mora and Rebeca Robles, National Institute of Psychiatry “Ramón de la Fuente”; Tahilia Rebello, Global Mental Health Program, Columbia University Medical Center, and New York State Psychiatric Institute; Pratap Sharan, Department of Psychiatry, All India Institute of Medical Sciences; Oye Gureje, Department of Psychiatry, University of Ibadan; Michael B. First and Howard F. Andrews, Department of Psychiatry, Columbia University Medical Center and New York State Psychiatric Institute; José Luís Ayuso-Mateos, Department of Psychiatry, Universidad Autónoma de Madrid; Wolfgang Gaebel and Juergen Zielasek, Department of Psychiatry, Heinrich-Heine University; Shekhar Saxena, Department of Mental Health and Substance Abuse, World Health Organization.

The authors of this article, with the exception of Geoffrey M. Reed and Shekhar Saxena, are members of or consultants to the WHO *ICD-11* Mental and Behavioural Disorders Field Study Coordination Group, reporting to the WHO International Advisory Group for the Revision of *ICD-10* Mental and Behavioural Disorders. Geoffrey M. Reed and Shekhar Saxena are members of the Secretariat, Department of Mental Health and

Substance Abuse, World Health Organization. The WHO Department of Mental Health and Substance Abuse has received direct support that contributed to the work described in this article from several sources: the International Union of Psychological Science, the National Institute of Mental Health (United States), the World Psychiatric Association, the Royal College of Psychiatrists (United Kingdom of Great Britain and Northern Ireland), the Spanish Foundation of Psychiatry and Mental Health (Spain), the Columbia University Global Mental Health Program (United States), the University of Kansas (United States), the Departments of Psychiatry, Universidad Autónoma de Madrid (Spain) and Universidad Nacional Autónoma de México (Mexico), and the Department of Psychology, University of Zurich (Switzerland). Unless specifically stated, the views expressed in this article are those of the authors and do not represent the official policies or positions of the WHO. The authors are grateful to Maya Kulygina, Valery Krasnov, and Anne M. Lovell for their helpful comments on an earlier version, and to the other participants in the Field Study Coordination Group—Brigitte Khoury, Jair de Jesus Mari, Toshimasa Maruta, Kathleen M. Pike, Min Zhao, Chihiro Matsumoto, Cary S. Kogan, and Anne-Claire Stona—for their contributions to the ideas presented in this article.

Correspondence concerning this article should be addressed to Geoffrey M. Reed, Department of Mental Health and Substance Abuse (MSD/MER), World Health Organization, 20, avenue Appia, CH-1211 Geneva 27, Switzerland. E-mail: [reedg@who.int](mailto:reedg@who.int)

*lems (ICD-11)*. The clinical utility of a diagnostic classification is critical to its function as the interface between health encounters and health information, and to making the *ICD-11* be a more effective tool for helping the WHO's 194 member countries, including the United States, reduce the global disease burden of mental disorders. This article describes the WHO's efforts to develop a science of clinical utility in regard to one of the two major classification systems for mental disorders. We present the rationale and methodologies for an integrated and complementary set of field study strategies, including large international surveys, formative field studies of the structure of clinicians' conceptualization of mental disorders, case-controlled field studies using experimental methodologies to evaluate the impact of proposed changes to the diagnostic guidelines on clinicians' diagnostic decision making, and ecological implementation field studies of clinical utility in the global settings in which the guidelines will ultimately be implemented. The results of these studies have already been used in making decisions about the structure and content of *ICD-11*. If clinical utility is indeed among the highest aims of diagnostic systems for mental disorders, as their developers routinely claim, future revision efforts should continue to build on these efforts.

*Keywords:* *ICD-11*, classification, clinical utility, field studies, *DSM-5*

The World Health Organization (WHO) is the world's foremost public health agency, whose mission is the attainment of the highest possible level of health by all people. The WHO's core responsibilities as a specialized agency of the United Nations are described in its constitution, ratified by international treaty with the WHO's 194 Member States, including the United States. Among the WHO's constitutional responsibilities are the development of international classification systems for health and the international standardization of diagnostic procedures (WHO, 2014b). This includes the *International Classification of Diseases and Related Health Problems (ICD)*, currently in its tenth revision (*ICD-10*; WHO, 1992a), used by WHO Member States as a mandatory framework for the collection and reporting of health statistics to the WHO. The WHO is currently undertaking a major revision of the *ICD*, and the *ICD-11* is expected to be presented to the World Health Assembly for approval in May 2018.

In addition to providing the international standard for health information, the *ICD* is increasingly used by WHO Member States as a framework for defining their responsibilities to provide free or subsidized health service to their citizens (International Advisory Group for the Revision of *ICD-10 Mental & Behavioral Disorders*, 2011). The *ICD* provides a common nomenclature for professionals and health care decision makers, a structure for organizing health care services, and an administrative framework for tracking, reimbursement, and outcomes assessment. As an integral part of the global classification for all health conditions, the *ICD* chapter on Mental and Behavioral Disorders is by far the most widely used classification of mental disorders worldwide (Reed, Mendonça Correia, Esparza, Saxena, & Maj, 2011). The first major revision of the *ICD* in more than two decades provides a major opportunity to improve the system, bringing it more in line with current

evidence and current practice. It is also a hugely complex undertaking with constituencies that include governments, health professionals, and users of health services and their families (International Advisory Group for the Revision of *ICD-10 Mental & Behavioral Disorders*, 2011).

The WHO's Department of Mental Health and Substance Abuse has responsibility for coordinating the technical activities involved in the development of the classification of Mental and Behavioral Disorders in *ICD-11*. In order to engage this work, the department appointed an International Advisory Group to provide overall conceptual guidance and a series of Working Groups with responsibility for reviewing the available scientific and clinical evidence and developing proposals for specific disorder areas. First, Reed, Hyman, and Saxena (2015) have provided a description of the process used for the development of diagnostic guidelines and of their specific content (see Maercker et al., 2013, and Tyrer, Reed, & Crawford, 2015, for examples of descriptions of proposals for specific disorder areas). The tasks of Working Groups included a consideration of the clinical utility and global applicability of proposals for the *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.; *DSM-5*; American Psychiatric Association, 2013), and an additional group was appointed to focus on harmonization between the two systems, with the goal of minimizing unintended and nonsubstantive differences.

Given the wide range of purposes for which the *ICD* is used and the dramatic variability in the supply of specialized mental health professionals by country income (WHO, 2014a), the WHO Department of Mental Health and Substance Use had produced three separate versions of the *ICD-10* classification of mental and behavioral disorders in addition to the main statistical version of the *ICD-10* covering all health conditions (WHO, 1992a). These included *Clinical Descriptions and Diagnostic Guidelines* intended

for use by mental health specialists in clinical settings (WHO, 1992b), a version for use by nonspecialists in primary care settings (WHO, 1996), and diagnostic criteria for research (WHO, 1993). Similarly, different presentations of the ICD-11 classification of mental and behavioral disorders are being developed for different groups of users (see First et al., 2015, for additional discussion of the differences among these versions and a detailed description of the material being developed for mental health specialists).

Over recent decades, the revision processes of both the ICD and the DSM have followed similar trajectories (Blashfield, Keeley, Flanagan, & Miles, 2014). A major focus of attention beginning in the 1970s was to improve the reliability of both systems through the introduction of a descriptive approach (Spitzer, Sheehy, & Endicott, 1977; WHO, 1974). It is widely acknowledged that the descriptive approach has markedly improved the identification and treatment of mental disorder throughout the world (Hyman, 2007), but this has likely not been through the application in everyday clinical practice of elaborate diagnostic criteria (Maj, 2015). Moreover, successive revisions of diagnostic manuals have not led to expected gains in the validity of the categories themselves (Cuthbert & Insel, 2013; Hyman, 2007, 2010). At the same time, it is overwhelmingly clear that there are major problems with the clinical utility of current classification systems—for example, excessive number of categories, overspecification, spurious comorbidity—that have grown worse over time (First, 2010; International Advisory Group for the Revision of ICD-10 Mental & Behavioral Disorders, 2011; Reed, 2010).

The purpose of this article is to describe a systematic and comprehensive program of field studies focused on clinical utility being conducted by the WHO Department of Mental Health and Substance Abuse as a part of the development of the ICD-11 classification of mental and behavioral disorders, conducted in collaboration with a Field Studies Coordination Group appointed by the WHO to provide consultation and scientific oversight for these studies. These field studies are substantially different from trials for previous versions of either the ICD or the DSM. They incorporate a variety of methods—many derived from psychology—with the specific goal of investigating and improving the clinical utility of the manual. The two main aims of the field studies program described in this article are (a) to provide *formative* information related to the nature and structure of diagnostic guidance for Mental and Behavioral Disorder to be provided in the ICD-11, and (b) to provide *evaluative* information regarding the performance of specific proposals developed by Working Groups in helping clinicians to reach diagnostic conclusions.

We begin by describing the WHO's conceptualization of clinical utility and discussing its importance for diagnostic classification systems. This is followed by a description of four field study strategies (two to address the *formative*

purpose; two to address the *evaluative* purpose) specifically designed to assess various aspects of clinical utility. Example results and conclusions for each type of completed study are provided to illustrate how the methodology addresses specific aspects of clinical utility and how the information derived from the studies is being used to develop the structure and content of the ICD-11 diagnostic manual. We conclude with a discussion of the limitations and differential contributions of the various methods and strategies being used as part of the ICD-11 field studies program, and how they can be used to improve future diagnostic manuals intended for implementation by clinicians as a part of health care encounters, and in developing a science of clinical utility.

### Definition and Importance of Clinical Utility

Based on earlier articulations of the concept (e.g., First et al., 2004), the WHO has defined the clinical utility of a classification as the ability of its diagnostic constructs to (a) facilitate communication among users of the system (professionals, researchers, patients, families, administrators, etc.), (b) facilitate conceptualization and understanding of mental disorders, (c) facilitate the system's implementation among those who will use it (e.g., how well the categories fit the patients' health providers see, how easy the system is to use and understand, how quickly clinicians reach a diagnostic conclusion), (d) help practitioners to select treatments or manage clinical conditions, and (e) improve clinical outcomes at the individual level and health status at the population level<sup>1</sup> (see Reed, 2010, Reed et al., 2013, and Roberts et al., 2012, for more complete discussions of the concept of clinical utility).

The WHO's focus on clinical utility in the current revision of the mental disorders classification relates to its mission as a global public health agency. Neuropsychiatric disorders account for approximately 13% of all global burden of disease and about one third of all disability (Whitford et al., 2013). Yet mental disorders are drastically undertreated, with between 35% and 50% of individuals with severe mental illness in high-income countries, and between 76% and 85% in low- and middle-income countries, receiving absolutely no treatment in the previous year (Demyttenaere et al., 2004). From the perspective of the WHO Department of Mental Health and Substance Abuse, a critical objective of the ICD-11 classification of mental and behavioral disorders is to serve as a more effective tool for helping WHO Member States to reduce the disease burden of these conditions. As noted by the International

<sup>1</sup> Mullins-Sweatt and Widiger (2009) have pointed out that it can be difficult to disentangle aspects of validity (particularly construct and predictive validity) from clinical utility. Rather than attempting to distinguish these constructs, we focus on the pragmatic application of the benefits of both validity and utility generally.

Advisory Group for the Revision of *ICD-10* Mental and Behavioral Disorders (2011), "People are only likely to have access to the most appropriate mental health services when the conditions that define eligibility and treatment selection are supported by a precise, valid, and clinically useful classification system" (p. 90).

Major gains in the clinical utility of the *ICD-11* classification of mental and behavioral disorders could have an important impact on public health. Classification systems provide the interface between health services and health information (Reed et al., 2013). A classification system that is not clinically useful will not be implemented as intended in clinical practice, and therefore cannot generate health encounter data that will be a valid basis for health statistics or the decisions that are based on them: eligibility determination, treatment management, and outcomes evaluation at the individual level; needs assessment, program development, and resource allocation at the system level; and public health decision-making at the national and global level.

The developers of modern classification systems routinely stress the importance of clinical utility as an orienting principle in their development. For example, the introduction to the *DSM-5* (American Psychiatric Association, 2013) indicates that clinical utility was one of the main considerations in making changes from the fourth edition of the *DSM* (*DSM-IV*; American Psychiatric Association, 1994), and responding to the specific needs of mental health clinicians was the rationale for developing the *Clinical Descriptions and Diagnostic Guidelines* for *ICD-10* Mental and Behavioral Disorders (WHO, 1992b). However, despite its stated importance, there has been virtually no systematic attention to the assessment of the clinical utility of classifications or its improvement in the development of modern classification systems. When it has been addressed at all as a part of field studies, it has been through brief subjective ratings by clinicians of ease of use, goodness of fit, and satisfaction (Clarke et al., 2013; Mościcki et al., 2013; Sartorius et al., 1993). Although clinician opinions are certainly one metric for clinical utility, they provide an insufficient scientific foundation for a meaningful conceptualization of clinical utility. Moreover, the methodologies applied to the study of this area have been quite weak given its universally endorsed importance. Clearly, a science of clinical utility has not yet been applied to the development of diagnostic classification systems.

### Clinical Utility Research Strategies for *ICD-11* Mental and Behavioral Disorders

No single study or type of study is sufficient to address all aspects of the clinical utility of a mental disorders classification; a multimethod, multi-informant, multilingual global research strategy is needed to generate hypotheses, inform

early decisions, and to test and improve proposals for diagnostic guidelines in *ICD-11*.

The goal of the initial, formative phase of the *ICD-11* field studies program was to provide guidance regarding the structure and scope of the classification before the development of diagnostic guidelines. This phase included two types of studies:

1. Professional surveys in collaboration with major international professional associations examined clinicians' perspectives about the classification of mental disorders, patterns of current use, and what changes they saw as necessary to improve the classification's clinical utility.
2. Formative field studies specifically designed to examine clinicians' cognitive organization of mental disorders were conducted in order to inform decisions about how the classification should be structured in order to provide a more intuitive interface with clinical practice.

The second, evaluative phase of the field studies program uses a different set of research strategies to evaluate clinicians' application of the proposed diagnostic guidelines. This phase includes the following types of studies:

1. Case-controlled studies use experimental methods in which the clinical case material is controlled through the use of standardized vignettes in order to evaluate the effects of the draft *ICD-11* diagnostic guidelines on diagnostic decision-making.
2. Ecological implementation studies examine the performance and usability of the proposed *ICD-11* diagnostic guidelines in naturalistic conditions with real patients, in settings around the world in which the guidelines are intended to be applied.

### Surveys of Health Professionals

Systematic surveys of professionals' use and perceived utility of diagnostic systems provide an important starting point for their revision, but previously available surveys have been limited because of small and highly selected samples, low response rate, or limited geographical scope (Mellsop et al., 2007; Suzuki et al., 2010; Zielasek et al., 2010). The WHO undertook two major surveys in collaboration with the World Psychiatric Association (WPA) and the International Union of Psychological Science (IUPsyS), the main international professional organizations representing psychiatrists and psychologists. Nearly 5,000 psychiatrists from 44 countries across all WHO global regions participated in the WHO-WPA survey in 19 languages

(Reed et al., 2011). Similarly, over 2,000 psychologists from 23 countries participated in the WHO-IUPsyS survey (Evans et al., 2013). These two surveys represent, by far, the largest and most geographically and linguistically diverse studies assessing the perceptions of psychologists and psychiatrists ever undertaken for the purpose of evaluating a diagnostic classification.

The large majority of global psychiatrists and the majority of global psychologists participating in these surveys primarily used the *ICD* in their day-to-day clinical practice, although professionals in a few countries, including the United States, primarily used the *DSM*. Respondents to both surveys indicated that the most important purposes of a classification of mental disorders were (a) to facilitate communication among clinicians, and (b) to inform treatment and management decisions. For use in daily clinical practice, both psychiatrists and psychologists overwhelmingly preferred a simpler classification with many fewer categories than current classification systems contain. A substantial majority of both groups of global clinicians preferred flexible guidelines that allow greater scope for cultural variation and clinical judgment compared with a strict criteria-based approach, regardless of whether they were users of the *ICD* or the *DSM*. Interestingly, this reflects a major difference between the approach taken by the two classifications.

Finally, approximately a third of participants from Asia, Europe, and Latin America, and slightly more in Africa and the Middle East, endorsed finding mental health diagnoses (*ICD-10* or *DSM-IV*) difficult to apply across cultures, compared with much smaller percentages in the United States. This finding suggests that more careful attention to cultural variation in the presentation of diagnostic guidelines is important. Almost half (46%) of participants found at least one diagnostic category especially problematic. The most problematic diagnostic grouping for both psychologists and psychiatrists was personality disorders. In addition, psychiatrists reported particular problems with schizoaffective and somatoform disorders, whereas psychologists were more likely to point to posttraumatic stress disorder (PTSD) and attention-deficit hyperactivity disorder as problematic (ADHD; Robles et al., 2014).

These findings have already guided the development of *ICD-11* in important ways. Over the past 35 years, the number of mental disorder categories in both the *ICD* and the *DSM* has dramatically expanded. The current *ICD* revision has particularly sought to eliminate or simplify categories or subtypes of limited clinical utility (or validity). Although new categories have been added, the clinical justification for their inclusion (e.g., differential treatment recommendations) has been carefully considered. Second, for classification systems intended to be used by practitioners, flexible guidelines are preferable to rigid, criteria-based rules. Criteria-based systems have been criticized as

inconsistent with clinical practice, the lack of evidence for specific symptom and duration requirements, and their oversimplification of complex psychopathological phenomena (Maj, 2015), which must be balanced against their ostensible benefits for reliability. The WHO has provided specific guidance related to the development of a consistent, guidelines-based system for *ICD-11* (First et al., 2015). Third, the WHO is addressing difficulties with the cross-cultural applicability of diagnostic guidelines in several ways, which are expected to reduce the perceived need for national adaptations of the *ICD-11*. All *ICD-11* Working Groups included members from all WHO regions and a high percentage of members from low- and middle-income countries (where the vast majority of the world's population lives). The WHO is making a major effort to field test all proposals globally and in multiple languages, and available information regarding cultural factors of each of the classification entities will be integrated into the *ICD-11* diagnostic guidelines. Fourth, specific categories that clinicians find to be problematic have been a particular focus of attention in developing the *ICD-11* diagnostic guidelines.

### Formative Field Studies

A set of initial studies was “formative” in the sense that these studies informed early decisions about the best architecture or structure for a classification of mental and behavioral disorders. Most available literature on this topic focuses on how the classification can best reflect the “true” relationships among mental disorders based on inconsistent and incomplete evidence using a range of possible validators (e.g., see Andrews et al., 2009; First, 2009; Hyman, 2010; Jablensky, 2009). However, a clinical utility perspective would suggest that if the function of a classification is to serve as an interface between health encounters and health information, the most important and desirable features of a classification's architecture would be that

(a) it helps clinicians find the categories that most accurately describe the patients they encounter as quickly, easily, and intuitively as possible and (b) the diagnostic categories so obtained would provide them with clinically useful information about treatment and management. (Reed et al., 2013, p. 1193).

A classification structure that is as consistent as possible with clinicians' cognitive structures of mental disorder categories is therefore most likely to be intuitive and easy to apply in clinical settings. The Field Studies Coordinating Group (FSCG) conducted two formative studies using very different methodologies to elucidate clinicians' conceptualizations of the structure of mental disorders.

The first study (Roberts et al., 2012) used a method of paired comparisons, a traditional psychometric technique that has been widely used in cognitive science, marketing, and psychological research, including previous studies of

how mental health professionals organize clinically relevant information (Egli, Schlatter, Streule, & Läge, 2006; Treat et al., 2002). The paired comparisons methodology is an implicit approach to investigating the cognitive dimensions used in making pairwise distinctions between stimuli. An advantage of the methodology is that it overcomes the tendency to apply overlearned, preexisting, explicit organizations—such as diagnostic classification systems—in making judgments of similarity among specific pairs of stimuli.

This study asked 1,371 psychiatrists and psychologists from 64 different countries, participating in English and Spanish, to rate the similarity of 100 pairs of disorders randomly selected from 1,170 possible combinations of 60 disorder categories selected to represent all *ICD-10* groupings of mental disorders. Roberts and colleagues (2012) then employed multidimensional scaling (MDS) analysis to “uncover” the implicit dimensions that clinicians were using to make similarity judgments and then to generate a cognitive map of mental disorders based on these dimensions. This step would be conceptually similar to generating a geographical map of a particular country based on knowledge of the distances between all towns in that country.

The analysis yielded a clear structure of three dimensions that accounted for clinicians’ views of the similarity among disorders. The first dimension reflected the internality versus externality of the disorder, with abuse of volatile solvents (inhalants) at one end and somatization disorder at the other. This dimension has also emerged in other structural studies of psychopathology (Krueger, Markon, Patrick & Iacono, 2005; Wright et al., 2013). The second dimension captured the developmental versus adult onset of the disorder, with conditions like autistic disorder at one end and substance-related disorders at the other. The third dimension represented the “functional” versus “organic” nature of the disorder, with separation anxiety disorder and pathological gambling at one end and vascular dementia and Alzheimer’s disease at the other. Although the dimensions themselves are directly recognizable and interpretable for most clinicians, plotting specific categories using these dimensions as axes yielded distinct groupings of disorders in three-dimensional space that roughly corresponded with the phenomenologically defined groups in current classification systems (e.g., psychotic disorders, substance use disorders, neurodevelopmental disorders).

The three-dimensional cognitive map of mental disorders produced the MDS analysis used in this study (see Roberts et al., 2012) was remarkably consistent across professions, languages, and WHO global regions ( $r_s \approx .80$ ). However, the organization of disorders produced by the three-dimensional cognitive map was not exactly the same as the organization of disorders in *ICD-10* or *DSM-IV* (the editions of the manuals in use when the study was conducted). The degree of similarity between the map of disorder categories based on clinicians’ similarity ratings and the orga-

nization of diagnostic systems was moderate ( $\kappa = .42$  for both *ICD-10* and *DSM-IV*). Interestingly, the proposed structure for *ICD-11* was a better fit ( $\kappa = .51$ ), more closely aligning with but not perfectly reflecting clinicians’ implicit understanding of the relationships among disorders.

This study suggests that cognitive structures held by clinicians are rational, highly stable, and shared, but depart in observable ways from existing classification systems. Clinicians use multiple dimensions simultaneously in thinking about a particular disorder. For example, in the cognitive map generated from clinician similarity ratings, social anxiety disorder is internalizing, developmental in origin, and “functional” compared with “organic,” whereas a hierarchical classification cannot account for three dimensions simultaneously. Clinicians also use additional information such as common comorbidities (e.g., between antisocial personality disorder and substance use disorders) in evaluating the distance between disorders. This information is clinically relevant, but is difficult to capture in the classification’s structure, except perhaps by placing related groups of disorders close to one another. The proposed structure for *ICD-11* contains fewer artificial conglomerations of disorders, and the order of the grouping in the classification is intended to be clinically meaningful rather than historical. Given the inherent limitations in the ability of a hierarchical classification to capture the complexity of clinicians’ conceptualizations, a moderate increase from *ICD-10* to *ICD-11* in the fit between the classification and the cognitive map generated by clinicians constitutes an important advance in the classification’s potential clinical utility.

Although the first formative field study was based on an implicit approach, the second formative study (Reed et al., 2013) investigated clinicians’ explicit constructions of the relationships among disorders. These two different methodological approaches complement each other, as both implicit and explicit representations will influence the way clinicians think about disorders and use a diagnostic manual. The second formative study was designed to investigate “natural taxonomies,” or the ways in which relevant phenomena are classified by specific cultures or populations (Berlin, 1992). Cognitive psychologists have developed systematic methodologies for understanding and measuring natural taxonomies among different groups of interest (Medin, Lynch, Coley, & Atran, 1997). Flanagan and colleagues adapted this methodology to study U.S. clinicians’ natural taxonomies of *DSM-IV* mental disorder categories (Flanagan & Blashfield, 2007; Flanagan, Keeley, & Blashfield, 2008, 2012). Reed and colleagues (2013) used the same basic methodology to investigate natural taxonomies of mental disorders among mental health professionals in eight countries (Brazil, China, India, Japan, Mexico, Nigeria, Spain, and the United States), in five different languages. Again, the idea was that if clinicians’ conceptualizations of the structure of mental disorder categories were

similar across countries, languages, and professions, this information could be used to make the organization of mental and behavioral disorders in *ICD-11* as clinically intuitive as possible.

A total of 517 mental health professionals, including at least 60 from each country, sorted index cards with the names of the same 60 disorders used in the first study into groups, and then aggregated and disaggregated their groupings to form a hierarchical organization of categories. Clinicians were instructed to form groups based on how they believed the categories should be organized to be as useful as possible in clinical practice in relationship to the identification and treatment of people with mental disorders, rather than attempting to replicate existing classifications. The hierarchical organizations produced by clinicians from different parts of the world, using different languages, and coming from different professions were highly consistent with each other ( $rs > .90$ ; see Reed et al., 2013). As with the first study, the results were not identical to either *ICD-10* or *DSM-IV*, so the consistency is not simply a product of familiarity with the prevailing diagnostic system. Rather, there were meaningful departures from each system. These differences were largely consistent with proposals for *ICD-11* (see Reed et al., 2013).

Again, the importance and surprising nature of the level of consistency across very different global populations of clinicians cannot be overstated. It suggests that the differences between existing classification systems and clinicians' cognitive structures can be meaningfully used in developing a new classification. One way to do so is by integrating aspects of clinicians' experience-based organization into the classification. For example, clinicians placed ADHD with neurodevelopmental disorders, and schizotypal disorder with other schizophrenia spectrum disorders, both of which are consistent with proposals for *ICD-11*. When clinician structures are not accommodated based on contradictory validity evidence, they can be used as a source of information about areas in which particular educational efforts are likely to be needed. For example, based on their grouping of categories in this study, clinicians did not perceive a relationship between obsessive-compulsive disorder and body dysmorphic disorder, both of which are included in a new proposed grouping of obsessive-compulsive and related disorders in *ICD-11*. Based in part on this finding, particular attention has been paid to articulating the clinical rationale for the grouping (e.g., common features of pathological repetitive unwanted thoughts and associated behaviors) in developing the diagnostic guidelines for this area (see Stein et al., 2016, for a detailed discussion).

### Evaluative Field Studies

As indicated, the second broad phase of the *ICD-11* field studies program focuses on the evaluation of specific proposals for *ICD-11* diagnostic guidelines developed by the

Working Groups. This phase includes both case-controlled studies and ecological implementation studies to examine how the diagnostic guidelines are likely to be used in clinical practice. The traditional method for investigating new mental disorder classifications is to provide practitioners with advance copies of the new classification, and ask them to conduct diagnostic assessments using the new system (e.g., Clarke et al., 2013). Specific methodologies may involve multiple ratings by the same professional (test-retest reliability) or single ratings by multiple professionals (interrater reliability), in addition to questions about how applicable the diagnostic guidelines were to the patient or how easy they were to use (clinical utility). Although such studies can be valuable for testing certain aspects of how the new system will be used in its intended setting, they are very limited in the extent to which they isolate the specific contribution of the new system to diagnostic conclusions. It is generally impossible to separate the contribution of the guidelines from variability in patients' presentations, even during separate interviews of the same patient. Further, these studies generally do not directly compare diagnostic conclusions reached using the new system with those using the old, so they cannot provide definitive conclusions regarding whether the new system represents an improvement. Finally, studies that focus on reliability as a primary outcome have often not been designed to examine the clinical decision-making process with sufficient granularity to determine the specific ways in which the guidelines or criteria could be improved when reliability is lower than would be desirable. The case-controlled field studies described in the next section address these issues by standardizing and manipulating the information that provides the basis for the clinician's diagnosis in order to isolate the effects of the classification system on diagnostic decision-making.

### Case-Controlled Field Studies

The case-controlled field studies for *ICD-11* currently being conducted by the WHO with the consultation of the FSCG employ standardized patient material in the form of clinical vignettes and a systematic diagnostic decision procedure that allows for the evaluation of the specific impact of changes in the classification from *ICD-10* to *ICD-11* on diagnostic decision making. In order to test the specific effect of the different guidelines, it is necessary to control for the variability associated with clinical presentations to which the guidelines would be applied and to manipulate key variables of interest. Vignette methodologies based on experimental designs are ideally and specifically suited to the examination of these research questions. Well-designed vignette studies retain both the external validity strengths of survey research and the internal validity strengths of experimental methods (see Evans et al., 2015). Moreover, re-

sponses of health professionals to vignette scenarios can be highly generalizable to “real world” decision making, showing greater reliability and validity than other methods without being subject to the ethical and feasibility limitations of conducting such research with real patients in clinical settings (e.g., Atzmüller & Steiner, 2010; Peabody, Luck, Glassman, Dresselhaus, & Lee, 2000; see Evans et al., 2015). Case-controlled studies using vignettes also have the enormous advantage that they can be conducted over the Internet to facilitate participation by large numbers of clinicians around the world.

**The Global Clinical Practice Network.** The WHO Department of Mental Health and Substance Abuse, via the FSCG, a range of professional societies, and its global network of Collaborating Centers has recruited a worldwide pool of health professionals called the Global Clinical Practice Network (GCPN) to serve as participants in case-controlled Internet-based field studies related to the development of the classification of mental and behavioral disorders in *ICD-11* (Reed et al., 2015). Any health professional who is legally authorized in her or his country to provide assessment or treatment services to individuals with mental and behavioral disorders is eligible to register online in any of nine languages (see [www.globalclinicalpractice.net](http://www.globalclinicalpractice.net)). Registrants provide extensive information about the nature of their background, experience, and practice that enables the selection of specifically targeted participant pools for case-controlled field studies. GCPN registrants agree to receive invitations to participate in Internet-based field studies no more than once a month, each requiring approximately 30 min of their time, though they may decline to participate in any particular study.

As of December 2015, the GCPN membership includes more than 12,300 clinicians from 144 countries. The two professional groups with the largest representation in the GCPN are physicians (53%), primarily psychiatrists, and psychologists (28%), and a range of other mental health professions such as psychiatric nursing, social work, and occupational therapy are also represented. All global regions are represented, and the regional distribution of the GCPN closely resembles the regional distribution of mental health professionals (WHO, 2014a). The professionals in the network represent a substantial amount of experience: The average clinical experience is 15.3 years ( $SD = 10.6$ ). Members of the GCPN must either supervise or directly provide clinical services to qualify for participation in most case-controlled field studies; 93% currently see patients, with 59% currently supervising the provision of services by other professionals. The network also represents professionals that work with a wide range of ages and specialties; this information allows the selection of samples that are specifically relevant for a given study. For example, GCPN registrants who indicate that they have special knowledge and experience in a particular area may be compared with reg-

istrants who do not. The participation of nonexperts and a broad range of disciplines is particularly important given the WHO’s goal of developing a classification system that is usable by a health professionals who may not have advanced training in the specific topic under study.

The GCPN participant pool is not a representative sample of all global clinicians, but rather a broadly international, multidisciplinary, and multilingual volunteer sample that is specifically interested in *ICD-11* field studies and is therefore likely to evaluate and apply the proposed diagnostic guidelines in a careful manner. The experimental manipulations used in the case-controlled field studies permit conclusions about the impact of the proposed *ICD-11* guidelines on clinical decision making (e.g., compared with *ICD-10*), and the GCPN membership is sufficiently large and diverse to permit meaningful comparisons among subgroups, such as by region, language, and profession. For example, if it is observed that clinicians participating in a particular study in Spanish apply a specific guideline differently than clinicians participating in English, this effect could reflect cultural or training differences or a translation problem. The causes of such differences can be further investigated and the guidelines adjusted accordingly and, if necessary, retested.

**General methodology for ICD-11 case-controlled field studies.** *ICD-11* case-controlled field studies use experimental designs, generally to test effects of differences between diagnostic guidelines for *ICD-10* and *ICD-11* on diagnostic decisions made by clinicians. After agreeing to participate in a particular study, clinicians are randomly assigned to review and use either the *ICD-10* or the *ICD-11* diagnostic guidelines in the specific area that is the focus of the study (e.g., disorders specifically associated with stress, feeding and eating disorders). Asking the same clinicians in a particular study to apply both the *ICD-10* and the *ICD-11* to the diagnosis of specific cases would create carryover and expectancy effects that would affect the result of the central comparison. Moreover, the central research question is not the clinicians’ comparative evaluation of the two sets of guidelines, but rather the consistency of their application to specific standardized cases. Therefore, the diagnostic system is manipulated as a between-participants variable. Participants are then presented with a series of two clinical vignettes (counterbalanced for order of presentation), selected from a small pool of paired-vignette comparisons carefully designed to investigate the effects of key changes proposed for *ICD-11*. For example, the introduction of a new required feature for a particular diagnosis in *ICD-11* compared with *ICD-10* would be tested using two vignettes that differ systematically in the presentation of that feature while remaining equivalent in other respects. The experimental methods test whether clinicians using *ICD-11* accurately identify and apply this new requirement and whether their diagnostic conclusions are different in the intended

ways from those of clinicians using *ICD-10*. As many planned comparisons are included in a particular study as necessary to evaluate the important differences between *ICD-11* and *ICD-10* in the area under study. Planned paired comparisons are most meaningful if the same clinicians evaluate both vignettes; therefore, this manipulation is within-participant.

For example, the proposed *ICD-11* diagnostic guidelines for PTSD include reexperiencing the traumatic event in the present as a diagnostic requirement (Maercker et al., 2013). The guideline is intended to mean that the individual experiences the traumatic event(s) as occurring again in the here and now, as opposed to reflecting on or ruminating about the event(s) and remembering the feelings experienced at the time. This requirement was not specified in the *ICD-10*. To test whether clinicians could accurately distinguish between reexperiencing and remembering as presented in clinical material, two vignettes were developed that manipulated this variable while including similar levels of other PTSD symptoms. This comparison was included as one of a series of eight paired comparisons included in the study on Disorders Specifically Associated with Stress (Keeley et al., 2015), each testing a specific change in the diagnostic guidelines and to which participating clinicians were randomly assigned. Participants assigned to the reexperiencing versus remembering comparison were asked to provide diagnoses for both vignettes. Clinicians using *ICD-10* should have assigned the diagnosis of PTSD to both vignettes, whereas clinicians using *ICD-11* should have assigned the PTSD diagnosis only to the vignette that included reexperiencing as a symptom.

After viewing each vignette, participants either select a diagnosis from the guidelines they reviewed or indicate that no diagnosis is warranted. (Subthreshold vignettes are used to evaluate the boundary between disorder and normal variation.) Subsequent questions are adaptively programmed on the Internet study platform to allow a step-by-step evaluation of the participant's decision-making process by reviewing the presence or absence of each diagnostic requirement for the disorder the clinician selected, after which participants are given an opportunity to change their diagnosis. Additional questions also assess the vignette (e.g., symptom severity, impairment, similarity to cases seen in clinical practice), the guidelines (e.g., ease of use, goodness of fit), and the clinicians' diagnostic response (e.g., confidence in diagnosis, differential diagnosis). When the participant has completed this process, it is repeated for the second vignette. The design enables causal inferences related to how differences between diagnostic systems and vignette material affect clinicians' diagnostic decision making, as well as the accuracy, efficiency, and clarity of the diagnosis. The step-by-step review of each diagnostic requirement for the diagnostic category assigned permits identification of specific aspects of the diagnostic guidelines that are inconsis-

tently evaluated by participating clinicians. Overall, these studies are a critical aspect of the *ICD-11* field studies program because they provide specific information regarding categories and diagnostic requirements that seem to be problematic, so that the corresponding aspects of the diagnostic guidelines can be improved and subsequently tested in additional case-controlled or ecological implementation field studies as a part of the current field studies program. Although information regarding participants' views of the diagnostic guidelines is collected, clinical utility in the case-controlled studies is experimentally operationalized as the extent to which the diagnostic guidelines help participants to reach accurate and consistent diagnostic conclusions in response to standardized case material.

**Vignette development and pretesting.** In the *ICD-11* case-controlled field studies, much of the weight of the study's validity is carried by the vignettes, which are therefore developed and tested according to a systematic and rigorous process (see Evans et al., 2015). Vignettes that manipulate specific variables to be used in each study are drafted by representatives of *ICD-11* Working Groups, who are global experts with substantial clinical and research experience in the area to be studied. Each vignette is developed according to general guidelines developed by the FSCG and specific instructions regarding the clinical characteristics required for each vignette specified in the study methodology. A different set of expert raters then pretests the vignettes to confirm the presence or absence of required diagnostic features, including the most appropriate diagnosis. This pretesting process ensures that departures by participants from the intended diagnostic conclusion represent changes in diagnostic reasoning rather than ambiguous content in the vignette.

Vignette developers and pretesters used in the *ICD-11* program of vignette-based field studies are always multidisciplinary and represent different world regions, including a high proportion from low- and middle-income countries and non-native English speakers. This process helps to ensure cultural neutrality of the vignettes and to reflect the intended users of the *ICD-11*. In addition, the vignette studies undergo rigorous processes for translation, retesting, and implementation in all languages, including Chinese, English, French, Japanese, Russian, and Spanish. The translation of case-controlled field studies into multiple languages has improved the cross-cultural applicability of the diagnostic guidelines, as phrases that are untranslatable or unclear in the English version are modified as a result of this process. This is a marked contrast to standard processes in which translation only occurs after the English text is finalized.

**Example result from a case-controlled field study.** As discussed previously, we described the *ICD-11* requirement of reexperiencing compared with remembering for a diagnosis of PTSD and how it was tested with paired vignettes. In the first case-controlled field study of disorders specifically asso-

ciated with stress, we found that many aspects of the proposed *ICD-11* guidelines were helpful in producing clearer and more consistent judgments among 1,738 GCPN registrants participating in the study in three languages (Keeley et al., 2015). However, participants did not make the distinction between reexperiencing and remembering as intended, as 68.8% of participants using *ICD-11* assigned to this comparison applied the PTSD diagnosis to the remembering vignette, compared with 81.3% for the reexperiencing vignette. Moreover, there was no difference between the participants who used the *ICD-11* and the *ICD-10* diagnostic guidelines as a basis for making this distinction. During the step-by-step review of diagnostic requirements, 38.6% of participants either indicated that they were aware that reexperiencing was not present in the remembering vignette or that they were not sure, but only 5.7% changed their initial diagnosis of PTSD when given the opportunity to do so. After reviewing these findings, the FSCG requested that the Working Group on Disorders Specifically Associated with Stress develop a clearer definition of reexperiencing for inclusion in the proposed diagnostic guidelines, and that the exclusion and appropriate diagnostic options under those circumstances be clarified. The revised guidelines will be used in subsequent studies. Although a negative finding, this example provides a useful illustration of how the *ICD-11* field studies are specifically intended to improve the clinical utility of the diagnostic guidelines.

### Ecological Implementation Field Studies

Although the case-controlled studies can support specific conclusions about the differential impact of *ICD-11* and *ICD-10* diagnostic guidelines on clinical decision making, it is still important to test the guidelines in real clinical settings in order to confirm that they indeed lead to improvements in the diagnostic process within the context they will be used—that is, to test their implementation “ecologically” in these settings. Therefore, the WHO plans to test the key changes for *ICD-11* diagnostic guidelines in a series of strategically designed ecological field studies to evaluate potential problems with clinical utility as well as the reliability of diagnostic conclusions in clinical settings.

As it is not possible within available time and resources to test the entire system, at this point, specific areas for ecological implementation field studies have been selected based on the conditions that represent the greatest proportion of global disease burden among mental disorders (Whiteford et al., 2013) and the highest levels of service utilization in mental health settings, as well as those areas in which potential implementation problems are suggested by the results of case-controlled field studies. Because these studies are considerably more expensive and labor-intensive than the Internet-based case-controlled field studies, they will focus on highly targeted questions using methodologies designed to provide specific evidence of the utility and reliability in clinical implementa-

tion. These methods have been selected to yield information that will be maximally useful in formulating the final versions of the diagnostic guidelines for inclusion in the published *ICD-11* manual.

The first arm of the ecological implementation studies is to evaluate the following aspects of the clinical utility of proposed *ICD-11* diagnostic guidelines: (a) their ability to aid clinicians’ understanding of the person’s condition, (b) how well they fit the presentation of actual clinical cases, (c) their feasibility of use in regular clinical interactions, and (d) their adequacy for assessing individual patients. The clinical utility arm of the ecological implementation studies will focus on new patients presenting to participating mental health treatment settings. The participating clinician will have been trained regarding the *ICD-11* diagnostic guidelines and will have them available during the patient’s assessment. After the clinical assessment is completed, the participating clinician will provide detailed information about the patient’s presentation and their diagnostic conclusions, including the step-by-step evaluation of the essential features of any diagnosis assigned and ratings of similar clinical utility parameters to those used in the case-controlled field studies. The use of common questions across the case-controlled and ecological implementation field studies will facilitate the consideration of similarities and differences and their relevance for the diagnostic guidelines.

The second arm of ecological implementation studies will evaluate the interrater reliability of diagnostic conclusions based on the proposed *ICD-11* diagnostic guidelines. The reliability arm will specifically target five groups of disorders: mood disorders, psychotic disorders, disorders specifically associated with stress, anxiety disorders, and common disorders of childhood. The reliability methodology will focus on the contribution of the diagnostic guidelines to interrater reliability, controlling for variance in clinical presentation or information available to the clinician. That is, the research question for the interrater reliability studies is whether two clinicians, based on the same information, agree on the diagnosis when using the diagnostic guidelines for *ICD-11*. Two clinicians who have not previously evaluated the enrolled patient will together conduct a diagnostic interview. Both clinicians will then independently provide a diagnostic assessment of the patient and will also provide a step-by-step evaluation of the essential features for each selected diagnosis (up to three) and the same clinical utility ratings used in the other studies. As for case-controlled studies, the step-by-step evaluation is important for determining specific sources of disagreement between pairs of clinicians.

Ecological implementation field studies will be implemented through the WHO’s network of International Field Study Centers in countries representing all WHO global regions and nearly 50% of the world’s population. Each International Field Study Center will coordinate a

network of local hospitals, clinics, or other facilities in its country or region, as needed, for particular studies. The clinical utility portion will be implemented in all participating centers to evaluate targeted diagnostic guidelines according the types of patients seeking care at each institution, and will also be open to participants in the GCPN. The reliability arm will be necessarily restricted to those centers with a sufficient number of clinicians and the available infrastructure to implement a relatively complex reliability protocol. In both arms of the ecological implementation studies, participating clinicians will use an electronic data platform that is similar to the one being used for the case-controlled field studies, which has proven to be a feasible tool for global data collection.

### Limitations of Each Method

No single type of study in the WHO's program of field studies for *ICD-11* mental and behavioral disorders can address all relevant questions related to clinical utility; each has its own limitations. The surveys conducted by the WHO in collaboration with the WPA and the IUPsyS provided information on the preferences and perceptions of psychiatrists and psychologists, including their evaluations of aspects of clinical utility (goodness of fit and ease of use) of diagnostic categories these professionals reported using at least once per week. However, the fact that the WHO solicited clinicians' opinions does not mean that these should override scientific evidence when this is available. Rather, systematic clinician reports based on their experience with mental disorders classification are regarded as an important source of information about areas of difficulty in implementation and the acceptability of the system to global practitioners. This information can inform decisions about the organization and content of the classification manual together with other forms of evidence. Similarly, the formative fields studies of clinicians' organization of mental disorders were not intended to provide evidence regarding the "real" nature, structure, and interrelationships among mental disorders, in contrast to other initiatives such as the National Institute of Mental Health's Research Do-

main Criteria (RDoC; Cuthbert & Insel, 2013). Rather, these empirical studies of clinicians' cognitive organization of diagnostic categories were intended to guide the WHO's efforts to build a maximally intuitive and clinically useful classification structure. The case-controlled field studies are being used to examine clinical decision making using the proposed diagnostic guidelines for *ICD-11* and whether the new system produces improvement compared with *ICD-11*. Their great strength is also their great weakness: The experimental control they provide separates them from the circumstances of their application in real clinical situations. However, this concern is balanced by the ecological implementation studies, which will examine the implementation of the *ICD-11* using real patients and under circumstances that are much more similar to health care practice, but that cannot in themselves separate other sources of diagnostic variability from the effects of the classification system being used.

### Conclusion

Like validity and reliability, the clinical utility of a diagnostic classification system is a multifaceted and complicated construct, but one that has been considerably less investigated. Also like validity and reliability, clinical utility is not a stable property of diagnostic guidelines, but rather depends on the specific purpose, context, and manner of their application. The strategy for field studies undertaken by the WHO related to the development of the classification of mental and behavioral disorders in *ICD-11* represents a serious effort at addressing these issues, employing multiple methodologies across multiple stakeholders from multiple cultures and world regions. These methods have ranged from opinion surveys to experimental designs, considering both internal and external study validity. As shown in Table 1, each of the methods being used provides coverage of different components of the WHO's definition of clinical utility. Although no study addresses every facet of clinical utility, in combination they provide adequate coverage of the construct. The WHO's aim is to gather a body of evi-

Table 1  
ICD-11 Field Study Coverage of WHO's Definition of Clinical Utility

Component of clinical utility	WPA & IUPsyS surveys	Formative field studies on classification structure	Case-controlled field studies	Ecological implementation field studies
Communication	✓	✓	✓	✓
Conceptualization		✓	✓	
Implementation	✓		✓	✓
Treatment/Management	✓			✓
Improve outcomes			✓	✓

Note. *ICD-11* = International Classification of Diseases and Related Health Problems, Eleventh Revision; WHO = World Health Organization; WPA = World Psychiatric Association; IUPsyS = International Union of Psychological Science.

dence across multiple domains that provides an empirical framework for the examination of clinical utility as part of the revision of diagnostic classifications and can be used as a basis for future efforts. The combination of methods being used balances the limitations of any one approach and, together, provide a much broader and deeper examination of clinical utility than efforts undertaken in the past.

Although the developers of previous diagnostic systems have claimed that clinical utility was a high priority, they have done relatively little to assess the utility of those systems or to weigh that information meaningfully as a part of efforts to improve them. Based on the state of current science, the developers of the *ICD-11* believe that only modest gains in the validity of the classification can be made as a part of the current revision. Equally, however, we believe that making meaningful progress on improving the *ICD-11*'s clinical utility is an attainable goal, and that this will lead to improvements in the identification and treatment of people who need mental health services.

The *ICD-11* field studies described in this article have already led directly to changes in the structure and content of the *ICD-11* diagnostic guidelines. Careful testing of how the diagnostic guidelines are applied by mental health professionals creates the opportunity to improve the manual in ways that are suggested by the results, supporting greater correspondence between clinicians' diagnostic decisions, the symptoms of their patients, and their treatment and clinical management choices. This process will continue up until the time that the new classification is approved by the World Health Assembly, anticipated in 2018.

We hope that the field trial process being undertaken in relation to the *ICD-11* classification of mental and behavioral disorders will set a new standard for the field. It is no longer sufficient in the face of global mental health challenges to assert that clinical utility is important; developers of diagnostic classifications must develop a science of clinical utility, using multiple empirical methods—many taken from psychology—to address specific aspects of the construct. Only in this way will diagnostic classification systems for mental disorders better meet the needs of people in need of mental health services, health professionals, governments, and other stakeholders, and better support the WHO's public health objective of reducing the disease burden of mental and behavioral disorders around the world.

## References

- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.
- Andrews, G., Goldberg, D. P., Krueger, R. F., Carpenter, W. T., Jr., Hyman, S. E., Sachdev, P., & Pine, D. S. (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. <http://dx.doi.org/10.1017/S0033291709990250>
- Atzmüller, C., & Steiner, P. M. (2010). Experimental vignette studies in survey research. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *6*, 128–138. <http://dx.doi.org/10.1027/1614-2241/a000014>
- Berlin, B. (1992). *Ethnobiological classification*. Princeton, NJ: Princeton University Press. <http://dx.doi.org/10.1515/9781400862597>
- Blashfield, R. K., Keeley, J. W., Flanagan, E. H., & Miles, S. R. (2014). The cycle of classification: *DSM-I* through *DSM-5*. *Annual Review of Clinical Psychology*, *10*, 25–51. <http://dx.doi.org/10.1146/annurev-clinpsy-032813-153639>
- Clarke, D. E., Narrow, W. E., Regier, D. A., Kuramoto, S. J., Kupfer, D. J., Kuhl, E. A., . . . Kraemer, H. C. (2013). *DSM-5* field trials in the United States and Canada, Part I: Study design, sampling strategy, implementation, and analytic approaches. *The American Journal of Psychiatry*, *170*, 43–58. <http://dx.doi.org/10.1176/appi.ajp.2012.12070998>
- Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: The seven pillars of RDoC. *BMC Medicine*, *11*, 126. <http://dx.doi.org/10.1186/1741-7015-11-126>
- Demyttenaere, K., Bruffaerts, R., Posada-Villa, J., Gasquet, I., Kovess, V., Lepine, J. P., . . . WHO 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. *JAMA: Journal of the American Medical Association*, *291*, 2581–2590. <http://dx.doi.org/10.1001/jama.291.21.2581>
- Egli, S., Schlatter, K., Streule, R., & Läge, D. (2006). A structure-based expert model of the *ICD-10* mental disorders. *Psychopathology*, *39*, 1–9. <http://dx.doi.org/10.1159/000089657>
- Evans, S. C., Reed, G. M., Roberts, M. C., Esparza, P., Watts, A. D., Correia, J. M., . . . Saxena, S. (2013). Psychologists' perspectives on the diagnostic classification of mental disorders: Results from the WHO-IUPsyS Global Survey. *International Journal of Psychology*, *48*, 177–193. <http://dx.doi.org/10.1080/00207594.2013.804189>
- Evans, S. C., Roberts, M. C., Keeley, J. W., Blossom, J. B., Amaro, C. M., Garcia, A. M., . . . Reed, G. M. (2015). Using vignette methodologies to study clinicians' decision-making behaviors: Validity, utility, recommendations, and application in field studies for *ICD-11*. *International Journal of Clinical and Health Psychology*, *15*, 160–170. <http://dx.doi.org/10.1016/j.ijchp.2014.12.001>
- First, M. B. (2009). Reorganizing the diagnostic groupings in *DSM-V* and *ICD-11*: A cost/benefit analysis. *Psychological Medicine*, *39*, 2091–2097. <http://dx.doi.org/10.1017/S0033291709991152>
- 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. <http://dx.doi.org/10.1037/a0021511>
- First, M. B., Pincus, H. A., Levine, J. B., Williams, J. B. W., Ustun, B., & Peele, R. (2004). Clinical utility as a criterion for revising psychiatric diagnoses. *The American Journal of Psychiatry*, *161*, 946–954. <http://dx.doi.org/10.1176/appi.ajp.161.6.946>
- First, M. B., Reed, G. M., Hyman, S. E., & Saxena, S. (2015). The development of the *ICD-11* Clinical Descriptions and Diagnostic Guidelines for Mental and Behavioural Disorders. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*, *14*, 82–90. <http://dx.doi.org/10.1002/wps.20189>
- Flanagan, E. H., & Blashfield, R. K. (2007). Clinicians' folk taxonomies of mental disorders. *Philosophy, Psychiatry, & Psychology*, *14*, 249–269. <http://dx.doi.org/10.1353/ppp.0.0123>
- Flanagan, E. H., 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. <http://dx.doi.org/10.1037/a0012535>

- Flanagan, E., Keeley, J., & Blashfield, R. (2012). Do clinicians conceptualize *DSM-IV* disorders hierarchically? *Journal of Clinical Psychology*, *68*, 620–630. <http://dx.doi.org/10.1002/jclp.21840>
- Hyman, S. E. (2007). Can neuroscience be integrated into the *DSM-V*? *Nature Reviews Neuroscience*, *8*, 725–732. <http://dx.doi.org/10.1038/nrn2218>
- Hyman, S. E. (2010). The diagnosis of mental disorders: The problem of reification. *Annual Review of Clinical Psychology*, *6*, 155–179. <http://dx.doi.org/10.1146/annurev.clinpsy.3.022806.091532>
- International Advisory Group for the Revision of ICD-10 Mental and Behavioural Disorders. (2011). A conceptual framework for the revision of the ICD-10 classification of mental and behavioural disorders. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*, *10*, 86–92. <http://dx.doi.org/10.1002/j.2051-5545.2011.tb00022.x>
- Jablensky, A. (2009). A meta-commentary on the proposal for a meta-structure for *DSM-V* and ICD-11. *Psychological Medicine*, *39*, 2099–2103. <http://dx.doi.org/10.1017/S0033291709991292>
- Keeley, J. W., Reed, G. M., Roberts, M. C., Evans, S. C., Robles, R., Matsumoto, C., . . . Maercker, A. (2015). Disorders specifically associated with stress: A case-controlled field study for *ICD-11* mental and behavioural disorders. *International Journal of Clinical and Health Psychology*. Advance online publication. <http://dx.doi.org/10.1016/j.ijchp.2015.09.002>
- Krueger, R. F., Markon, K. E., Patrick, C. J., & Iacono, W. G. (2005). Externalizing psychopathology in adulthood: A dimensional-spectrum conceptualization and its implications for *DSM-V*. *Journal of Abnormal Psychology*, *114*, 537–550. <http://dx.doi.org/10.1037/0021-843X.114.4.537>
- Maercker, A., Brewin, C. R., Bryant, R. A., Cloitre, M., van Ommeren, M., Jones, L. M., . . . Reed, G. M. (2013). Diagnosis and classification of disorders specifically associated with stress: Proposals for ICD-11. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*, *12*, 198–206. <http://dx.doi.org/10.1002/wps.20057>
- Maj, M. (2015). The media campaign on the *DSM-5*: Recurring comments and lessons for the future of diagnosis in psychiatric practice. *Epidemiology and Psychiatric Sciences*, *24*, 197–202. <http://dx.doi.org/10.1017/S2045796014000572>
- Medin, D. L., Lynch, E. B., Coley, J. D., & Atran, S. (1997). Categorization and reasoning among tree experts: Do all roads lead to Rome? *Cognitive Psychology*, *32*, 49–96. <http://dx.doi.org/10.1006/cogp.1997.0645>
- Mellso, G., Banzato, C., Shinfuku, N., Nagamine, M., Pereira, M., & Dutu, G. (2007). An international study of the views of psychiatrists on present and preferred characteristics of classifications of psychiatric disorders. *International Journal of Mental Health*, *36*, 17–25. <http://dx.doi.org/10.2753/IMH0020-7411360402>
- Mościcki, E. K., Clarke, D. E., Kuramoto, S. J., Kraemer, H. C., Narrow, W. E., Kupfer, D. J., & Regier, D. A. (2013). Testing *DSM-5* in routine clinical practice settings: Feasibility and clinical utility. *Psychiatric Services*, *64*, 952–960. <http://dx.doi.org/10.1176/appi.ps.201300098>
- Mullins-Sweatt, S. N., & Widiger, T. A. (2009). Clinical utility and *DSM-V*. *Psychological Assessment*, *21*, 302–312. <http://dx.doi.org/10.1037/a0016607>
- Peabody, J. W., Luck, J., Glassman, P., Dresselhaus, T. R., & Lee, M. (2000). Comparison of vignettes, standardized patients, and chart abstraction: A prospective validation study of 3 methods for measuring quality. *JAMA: Journal of the American Medical Association*, *283*, 1715–1722. <http://dx.doi.org/10.1001/jama.283.13.1715>
- Reed, G. M. (2010). Toward ICD-11: Improving the clinical utility of WHO's international classification of mental disorders. *Professional Psychology: Research and Practice*, *41*, 457–464. <http://dx.doi.org/10.1037/a0021701>
- Reed, G. M., Mendonça Correia, J., Esparza, P., Saxena, S., & Maj, M. (2011). The WPA-WHO Global Survey of Psychiatrists' Attitudes Toward Mental Disorders Classification. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*, *10*, 118–131. <http://dx.doi.org/10.1002/j.2051-5545.2011.tb00034.x>
- Reed, G. M., Rebello, T. J., Pike, K. M., Medina-Mora, M. E., Gureje, O., Zhao, M., . . . Saxena, S. (2015). WHO's Global Clinical Practice Network for mental health. *Lancet Psychiatry*, *2*, 379–380. [http://dx.doi.org/10.1016/S2215-0366\(15\)00183-2](http://dx.doi.org/10.1016/S2215-0366(15)00183-2)
- Reed, G. M., Roberts, M. C., Keeley, J., Hoopell, C., Matsumoto, C., Sharan, P., . . . Medina-Mora, M. E. (2013). Mental health professionals' natural taxonomies of mental disorders: Implications for the clinical utility of the ICD-11 and the *DSM-5*. *Journal of Clinical Psychology*, *69*, 1191–1212. <http://dx.doi.org/10.1002/jclp.22031>
- Roberts, M. C., Reed, G. M., Medina-Mora, M. E., Keeley, J. W., Sharan, P., Johnson, D. K., . . . Saxena, S. (2012). A global clinicians' map of mental disorders to improve ICD-11: Analysing meta-structure to enhance clinical utility. *International Review of Psychiatry*, *24*, 578–590. <http://dx.doi.org/10.3109/09540261.2012.736368>
- Robles, R., Fresan, A., Evans, S. C., Lovell, A., Medina-Mora, M. E., Maj, M., & Reed, G. M. (2014). Problematic, absent and stigmatizing diagnoses in current mental disorders classifications: Results from the WHO-WPA and WHO-IUPsYs Global Surveys. *International Journal of Clinical and Health Psychology*, *14*, 165–177. <http://dx.doi.org/10.1016/j.ijchp.2014.03.003>
- Sartorius, N., Kaelber, C. T., Cooper, J. E., Roper, M. T., Rae, D. S., Gulbinat, W., . . . 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*, 115–124. <http://dx.doi.org/10.1001/archpsyc.1993.01820140037004>
- Spitzer, R., Sheehy, M., & Endicott, J. (1977). *DSM-III: Guiding principles*. In V. Rakoff, H. Stancer, & H. Kedward (Eds.), *Psychiatric diagnosis* (pp. 1–24). New York, NY: Brunner/Mazel.
- Stein, D. J., Kogan, C. S., Atmaca, M., Fineberg, N. A., Fontenelle, L. F., Grant, J., . . . Reed, G. M. (2016). The classification of obsessive-compulsive and related disorders in the *ICD-11*. *Journal of Affective Disorders*, *190*, 663–674. <http://dx.doi.org/10.1016/j.jad.2015.10.061>
- Suzuki, Y., Takahashi, T., Nagamine, M., Zou, Y., Cui, J., Han, B., . . . Shinfuku, N. (2010). Comparison of psychiatrists' views on classification of mental disorders in four East Asian countries/area. *Asian Journal of Psychiatry*, *3*, 20–25. <http://dx.doi.org/10.1016/j.ajp.2009.12.004>
- Treat, T. A., McFall, R. M., Viken, R. J., Nosofsky, R. M., MacKay, D. B., & Kruschke, J. K. (2002). Assessing clinically relevant perceptual organization with multidimensional scaling techniques. *Psychological Assessment*, *14*, 239–252.
- Tyrer, P., Reed, G. M., & Crawford, M. J. (2015). Classification, assessment, prevalence, and effect of personality disorder. *The Lancet*, *385*, 717–726. [http://dx.doi.org/10.1016/S0140-6736\(14\)61995-4](http://dx.doi.org/10.1016/S0140-6736(14)61995-4)
- Whiteford, H. A., Degenhardt, L., Rehm, J., Baxter, A. J., Ferrari, A. J., Erskine, H. E., . . . Vos, T. (2013). Global burden of disease attributable to mental and substance use disorders: Findings from the Global Burden of Disease Study 2010. *The Lancet*, *382*, 1575–1586. [http://dx.doi.org/10.1016/S0140-6736\(13\)61611-6](http://dx.doi.org/10.1016/S0140-6736(13)61611-6)
- World Health Organization. (1974). *Glossary of mental disorders and guide to their classification*. Geneva, Switzerland: Author.
- World Health Organization. (1992a). *The international classification of diseases and related health problems* (10th Revision). Geneva, Switzerland: Author.
- World Health Organization. (1992b). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines*. Geneva, Switzerland: Author. Retrieved from <http://www.who.int/classifications/icd/en/bluebook.pdf>
- World Health Organization. (1993). *The ICD-10 classification of mental and behavioural disorders: Diagnostic criteria for research*. Geneva, Switzerland: Author. Retrieved from <http://apps.who.int/iris/bitstream/10665/37108/1/9241544554.pdf?ua=1&ua=1>

- World Health Organization. (1996). *Diagnostic and management guidelines for mental disorders in primary care: ICD-10 Chapter V Primary Care Version*. Göttingen, Germany: Hogrefe and Huber.
- World Health Organization. (2014a). *Mental health atlas 2014*. Geneva, Switzerland: Author. Retrieved from [http://www.who.int/mental\\_health/evidence/atlas/mental\\_health\\_atlas\\_2014/en/](http://www.who.int/mental_health/evidence/atlas/mental_health_atlas_2014/en/)
- World Health Organization. (2014b). Constitution of the World Health Organization. In *World Health Organization basic documents* (48th ed., pp. 7–25). Geneva, Switzerland: Author. Retrieved from <http://apps.who.int/gb/bd/PDF/bd48/basic-documents-48th-ed.-en.pdf>
- Wright, A. G., Krueger, R. F., Hobbs, M. J., Markon, K. E., Eaton, N. R., & Slade, T. (2013). The structure of psychopathology: Toward an expanded quantitative empirical model. *Journal of Abnormal Psychology, 122*, 281–294. <http://dx.doi.org/10.1037/a0030133>
- Zielasek, J., Freyberger, H. J., Jänner, M., Kapfhammer, H. P., Sartorius, N., Stieglitz, R. D., & Gaebel, W. (2010). Assessing the opinions and experiences of German-speaking psychiatrists regarding necessary changes for the 11th revision of the mental disorders chapter of the International Classification of Disorders (ICD-11). *European Psychiatry, 25*, 437–442. <http://dx.doi.org/10.1016/j.eurpsy.2009.12.023>

Received November 21, 2014

Revision received April 28, 2015

Accepted July 13, 2015 ■