Mental Health Professionals' Natural Taxonomies of Mental Disorders: Implications for the Clinical Utility of the ICD-11 and the DSM-5

Geoffrey M. Reed,[1] Michael C. Roberts,[2] Jared Keeley,[3] Catherine Hooppell,[1] Chihiro Matsumoto,[4] Pratap Sharan,[5] Rebeca Robles,[6] Hudson Carvalho,[7] Chunyan Wu,[8] Oye Gureje,[9] Itzear Leal-Leturia,[10] Elizabeth H. Flanagan,[11] João Mendonça Correia,[1] Toshimasa Maruta,[4] José Luís Ayuso-Mateos,[10] Jair de Jesus Mari,[7] Zeping Xiao,[12] Spencer C. Evans,[2] Shekhar Saxena,[1] and María Elena Medina-Mora[6]

[1] *World Health Organization*
[2] *University of Kansas*
[3] *Mississippi State University*
[4] *Tokyo Medical University*
[5] *All India Institute of Medical Sciences*
[6] *National Institute of Psychiatry Ramon de la Fuente Muñiz*
[7] *Federal University of São Paulo*
[8] *Shanghai Mental Health Center*
[9] *University of Ibadan*
[10] *Universidad Autónoma de Madrid*
[11] *Yale School of Medicine*
[12] *Shanghai Jiaotong University School of Medicine*

Objective: To examine the conceptualizations held by psychiatrists and psychologists around the world of the relationships among mental disorders in order to inform decisions about the structure of the classification of mental and behavioral disorders in World Health Organization's International Classification of Diseases and Related Health Problems 11th Revision (ICD-11). **Method:** 517 mental health professionals in 8 countries sorted 60 cards containing the names of mental disorders into groups of similar disorders, and then formed a hierarchical structure by aggregating and disaggregating these groupings. Distance matrices were created from the sorting data and used in cluster and

The authors are grateful to Dr. Roger Blashfield for his very helpful comments regarding the design of the study and the interpretation and presentation of the data. The World Health Organization Department of Mental Health and Substance Abuse has received direct support that contributed to the conduct of this study from several sources: The International Union of Psychological Science, the National Institute of Mental Health (USA), the World Psychiatric Association, the Spanish Foundation of Psychiatry and Mental Health (Spain), and the Santander Bank UAM/UNAM endowed Chair for Psychiatry (Spain/Mexico). P. Sharan received a grant from Eli Lilly (India) unrelated to this project; J. Mari received speaking fees from Astra Zeneca, Eli Lilly, and Janssen unrelated to this project.

Most of the authors of this article are members of the WHO International Advisory Group for the Revision of ICD-10 Mental and Behavioural Disorders and/or of Working Groups that report to the International Advisory Group. G. Reed and S. Saxena are members of the WHO Secretariat, Department of Mental Health and Substance Abuse, WHO. The views expressed in this article are those of the authors and, except as specifically noted, do not represent the official policies or positions of the International Advisory Group or of WHO. The WHO Research Ethics Review Committee as well as applicable local Institutional Review Boards reviewed and approved all procedures.

João Correia is now at the Department of Cognitive Neuroscience, Maastricht University, Maastricht, The Netherlands.

Please address correspondence to: Dr. Geoffrey Reed, Department of Mental Health and Substance Abuse, World Health Organization, CH-1211 Geneva 27, Switzerland. E-mail: reedg@who.int

JOURNAL OF CLINICAL PSYCHOLOGY, Vol. 69(12), 1191–1212 (2013)

Published online in Wiley Online Library (wileyonlinelibrary.com/journal/jclp). DOI: 10.1002/jclp.22031

correlation analyses. **Results:** Clinicians' taxonomies were rational, interpretable, and extremely stable across countries, diagnostic system used, and profession. Clinicians' consensus classification structure was different from ICD-10 and the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders 4th Edition (DSM-IV), but in many respects consistent with ICD-11 proposals. **Conclusions:** The clinical utility of the ICD-11 may be improved by making its structure more compatible with the common conceptual organization of mental disorders observed across diverse global clinicians. J. Clin. Psychol. 69:1191–1212, 2013.

Keywords: mental disorders; classification, International Classification of Diseases (ICD); Diagnostic and Statistical Manual of Mental Disorders (DSM); clinical utility; cross-cultural applicability; low- and middle-income (LAMI) countries

The World Health Organization (WHO) is currently revising the International Classification of Diseases and Related Health Problems Tenth Revision (ICD-10; WHO, 1992), with the ICD-11 slated for approval by the World Health Assembly in 2015. The WHO Department of Mental Health and Substance Abuse is responsible for managing the technical work of developing the ICD-11 chapter on mental and behavioral disorders, within the context of broader policies developed by the overall ICD-11 classifications team. Serious problems with the clinical utility of both the ICD-10 and the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders 4th Edition (DSM-IV) classifications of mental disorders are widely acknowledged (e.g., Andrews et al., 2009; Kendell & Jablensky, 2003; First, 2010). Some of these problems are: (a) extensive use of "Unspecified" or "Not Otherwise Specified" categories of no informational value; (b) artificial and inflated comorbidity among mental disorders categories; (c) many of the distinctions clinicians are asked to make in diagnostic classification systems have no relevance for treatment, while important diagnostic heterogeneity in other areas is obscured; and (d) the sheer complexity of current diagnostic systems, with each revision including more categories, more subtypes, and more specifiers (see Reed, 2010).

A major goal of the WHO Department of Mental Health and Substance Abuse for the current revision is to improve the clinical utility of this part of the ICD-11 (Reed, 2010; International Advisory Group for the Revision of ICD-10 Mental and Behavioural Disorders, 2011). "People are only likely to have access to the most appropriate mental health services when the conditions that define identification, eligibility and treatment selection are supported by a precise, valid, and clinically useful classification system" (International Advisory Group for the Revision of ICD-10 Mental and Behavioural Disorders, 2011, p. 90). In order for the ICD-11 classification of mental and behavioral disorders to be a more effective tool for meeting international public health goals, the new system will need to be usable for implementation throughout the world at the point where people with mental health needs are most likely to come into contact with the health system.

What Is Clinical Utility and Why Is It Important to WHO?

For the purpose of this program of work and drawing on previous definitions of clinical utility (First et al., 2004; Kendell & Jablensky, 2003; Mullins-Sweatt & Widiger, 2009), WHO has conceptualized the clinical utility of a classification, construct, or category for mental and behavioral disorders as: (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); (c) its usefulness in *selecting interventions* and making *clinical management* decisions; and (d) the extent to which it is associated with improvements in clinical outcomes at the individual level and in health status at the population level (see Reed, 2010).

WHO is interested in clinical utility because it is critical to the interface between clinical practice and health information. Global health care systems are overburdened; clinicians are under enormous time pressure, and only a very small minority of persons with mental health needs will ever see a specialist mental health professional. A mental disorders classification that

is difficult and cumbersome to implement in clinical practice and does not provide information that is of immediate value to the clinician has no hope of being implemented accurately at the encounter level in real-word health care settings (Reed, 2010; Roberts et al., 2012). In that event, clinical practice will not be guided by the standardization and operationalization of concepts and categories that are inherent in the classification, and important opportunities for practice improvement and outcomes assessment will be lost. In turn, a diagnostic system that is characterized by poor clinical utility at the encounter level cannot generate data based on those encounters that will be a valid basis for health programs and policies, or for global health statistics.

Clinical Utility and the Architecture of Diagnostic Classification

The appropriate architecture of a diagnostic classification of mental and behavioral disorders is an issue that has received substantial recent attention. Andrews et al. (2009) offered a particularly ambitious proposal of five broad clusters of mental and behavioral disorders and examined the evidence for these a priori clusters based on eleven "validators." The authors claimed that such a simplified structure would enhance clinical utility by reducing complexity. However, this reduction in complexity was partly due to the fact that more than half of existing mental disorders categories were not encompassed by the five groupings (First, 2009). The authors did not explain the mechanisms through which an improvement in clinical utility would occur or provide any supporting evidence for this claim, which was strenuously challenged in a series of commentaries published simultaneously with the proposals (First, 2009; Jablensky, 2009; Wittchen, Beesdo, & Gloster, 2009).

Andrews and colleagues' implicit perspective is that what is most important in creating the structure of a mental disorder classification is to reflect the most current and accurate scientific understanding about the "true" relationships among mental disorders. Unfortunately, as Jablensky (2009) and Wittchen et al. (2009) pointed out, a review of currently available evidence does not provide uniform or definitive support for one particular architecture. Moreover, Hyman (2010) has suggested that the foundation for such an effort is shaky when many of the disorders themselves may not be valid and distinct disease entities; it is not possible to have a valid biomarker for a fictive category.

From a clinical utility perspective, particularly in terms of improving the interface between health information and clinical practice, the most important and desirable features of a classification's organization 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. An organization based on Andrews et al.'s (2009) "validators" would not contribute to this objective unless individual data for these parameters (e.g., genetic risk factors, biomarkers) were routinely available in the clinical setting and contributed meaningfully to treatment and management decisions, which for many of them is clearly not currently the case.

Given current technology, it is possible to develop a comprehensive and uniform taxonomic system that underlies a classification, with unique identifiers for each individual category to facilitate interoperability across different applications that are distinct from how those categories are presented to users. For example, the structure and logic of airline reservation codes are meaningless to most travelers—who are more concerned with the date, time, and destination of their flights—but perfectly interpretable to computers. The underlying architecture of the identifiers is completely independent of how that information might be presented to users for a particular purpose. As an illustration of a similar process, different Amazon.com users can find the same book by searching under mysteries, bestsellers, the works of a particular author, or books published in June 2011, rather than having to use the 10-digit International Standard Book Number (ISBN) of the International Organization for Standardization (ISO) to make their purchase; that is, the unique identifier for each book—its code in the classification—is independent of how the material is presented to the user, because the presentation can be

organized by any attribute that is part of the underlying taxonomy (e.g., author, theme, date of publication).

Parallel electronic characteristics are central to the current ICD-11 development process across all disease areas, of which the mental and behavioral disorders chapter is only a part. Therefore, the question of what is the best architecture for the classification can be framed as an inquiry into what organization will be most useful for the system's intended users in a particular context (e.g., use in tertiary mental health settings, primary care, research, public health), rather than as an argument about the single and most correct way to summarize inconsistent and incomplete evidence about the "true" relationships among mental disorder categories.

It was in this context that the current study set out to explore the subjective classification systems of mental disorders held by mental health professionals. The purpose of this study was to examine the conceptualizations held by psychiatrists and psychologists around the world of how mental disorders are related to one another, to use that information together with available validity evidence in making decisions about the "metastructure" of the ICD-11 classification of mental and behavioral disorders. If clinicians' "natural taxonomies" are consistent across countries, languages, and disciplines—and also distinct from current formal classification systems—then this information could be used to create a classification of mental disorders that corresponds more closely to clinicians' cognitive organization of categories and is therefore more intuitive and efficient for use in real-world health care settings. This could contribute to better implementation of the classification and improved health encounter data quality and, in turn, help to make the ICD-11 a better tool in relationship to WHO's public health priority of improving the identification and treatment of mental disorders at the country level (WHO, 2012).

Clinicians' Natural Taxonomies of Mental Disorders

If one considers the corpus of mental health professionals as representing a type of "culture" in that they share common beliefs, experiences, and purposes, then a model exists for understanding and measuring the ways in which they classify the phenomena most relevant to them (Flanagan & Blashfield, 2007). "Natural" or "folk" taxonomies have long been studied by cultural anthropologists because of what these reveal about how cultures organize and categorize the living world around them (Berlin, 1992). Often cultures develop detailed and complex classifications for living creatures or environmental phenomena because this knowledge serves a useful purpose (e.g., which plants are good to eat and which are poisonous). Cognitive psychologists have developed these ideas into a systematic methodology for studying natural taxonomies among a variety of different groups (Medin, Lynch, Coley, & Atran, 1997).

Mental health professionals inherently organize the salient aspects of their environment (i.e., individuals seeking treatment) to facilitate a variety of purposes, such as communicating with other professionals, establishing prognoses, facilitating access to information, and making treatment decisions (Blashfield, Keeley, & Burgess, 2009). Some of this inherent classificatory activity has been systematized over time in theories of psychopathology as well as in formal classification systems such as the ICD and the DSM. However, these taxonomies also exist at an informal level in the cognitive, conceptual structures of individual clinicians (Cantor, Smith, French, & Mezzich, 1980).

In a series of studies, Flanagan and colleagues (Flanagan & Blashfield, 2006, 2007; Flanagan, Keeley, & Blashfield, 2008, 2012) applied cognitive psychological methods to investigating the natural taxonomies of mental disorders held by American psychologists and psychiatrists. Participants were asked to sort index cards with the names of mental disorders into groups. The clinicians were instructed to put the disorders together based upon how they felt they should go together, rather than simply replicating the structure of the DSM or another classification system. The resulting structure provides a glimpse into the cognitive organization of individual clinicians, as well as a measure of how similar or divergent professionals were. These studies were conceptually interesting and potentially relevant to the ICD revision process, but were small and geographically narrow in scope. We sought to expand this examination to a more global level

that could provide more definitive information that would be useful for the development of the ICD-11.

The Present Study

The present study was conducted as a project of the Field Studies Coordination Group for ICD-11 Mental and Behavioural Disorders during the formative phases of the ICD revision, as part of a broader program of surveys and field studies intended to inform early decisions about the basic structure and content of the classification. The primary purpose of the study was to elucidate the "natural taxonomies" of mental and behavioral disorders that mental health professionals used in their day-to-day practice. Based on that information, the study sought to examine whether clinicians' conceptual structures for classifying mental disorders appeared to be common across global mental health professionals or differed according to country, language, or profession. Finally, the study aimed to determine the extent to which clinicians' taxonomies replicated existing classification systems or otherwise varied according to the classification system they actually used (e.g., ICD, DSM).

More than 500 mental health professionals, recruited through International Field Study Centers in eight countries and five continents, participated in the present study. This mix of participants was designed to include individuals from multiple regions, languages, and cultures—including non-Western societies—to allow an examination of whether these variables influence clinicians' classifications. Participation involved an individually administered session of 60 to 90 minutes with a trained experimenter, conducted in the local language of that country, in which participants were asked to generate a classification of mental disorder categories based on their own clinical experience. To be eligible for the study, participants had to spend a minimum of 10 hours per week providing direct mental health services to patients. In addition, participants had to have a minimum of 2 years of professional experience posttraining. This was based on evidence that during training mental health professionals' conceptual structure of mental disorders more closely resembles the formal classification systems they are being taught and that a more differentiated "expert" model emerges over time (Egli, Schlatter, Streule, & Läge, 2006; Egli, Streule, & Läge, 2008).

The methodology for the study was partly based on the previous U.S. study of clinicians' "natural" taxonomies of mental disorders described above (Flanagan et al., 2008), with some important innovations. As in the earlier study, clinicians were presented with a set of cards, each containing the name of a mental disorder, and were asked to sort the cards into groups based on their own clinical experience and how they approached the clinical management of these conditions, without attempting to replicate the structure of the ICD or DSM. They were then asked to perform aggregations and disaggregations of their groupings to illuminate each participant's implicit hierarchy. In the Flanagan et al. (2008) study, an attempt was made to elicit each participant's entire hierarchical structure, however many levels that might be. However, a feasible formal classification of mental disorder can only realistically contain a certain number of hierarchical levels of categories. Based on how participants in the Flanagan et al. study had approached the hierarchical aspect of the task, we designed the procedure to elicit a maximum of three hierarchical levels from each participant, in an effort to limit its complexity and time requirements.

Method

Participants

The sample comprised 517 mental health professionals recruited through eight WHO-appointed International Field Study Centers in Brazil (Federal University of São Paulo), People's Republic of China (Shanghai Mental Health Center), India (All India Institute of Medical Sciences), Japan (Tokyo Medical University), Mexico (National Institute of Psychiatry Ramón de la Fuente Muñiz), Nigeria (University of Ibadan), Spain (Universidad Autónoma de Madrid), and the United States (University of Kansas). These Centers are working officially with WHO as a part

of the ICD revision and were selected because of WHO's confidence in their capacity to conduct the study, their willingness to participate, and their location in widely diverse countries and regions. At least 60 participants were recruited per International Field Study Center (as shown in Table 2).

Procedures

All procedures for the study had been approved by the WHO Research Ethics Review Committee and appropriate local institutional review boards for the International Field Study Centers.

Recruitment. Individuals who met the following eligibility criteria were recruited at each Center: licensed or authorized mental health professionals with more than 2 years experience posttraining and whose current professional activities included providing mental health services to patients (i.e., direct patient contact involving assessment, treatment, or management) for at least 10 hours per week.

Experimenters and training. Experimenters for this study were either graduate-level research assistants with a background in mental health or doctoral-level co-investigators. Each experimenter participated in a 3-hour training session conducted by either the first author or one of the co-investigators. Experimenters were provided with all necessary materials and a detailed script to follow throughout all phases of the study. Posttraining follow-up procedures ensured that the experimental sessions were implemented consistently and correctly. Data entered by experimenters directly into the study website were checked for protocol compliance and accuracy on a continuous basis.

Languages. All materials were provided and all procedures conducted in the local language of the country in which the International Field Study Center was located. Languages included Chinese, English, Japanese, Portuguese, and Spanish. Translations of all materials from English were prepared under the supervision of the International Field Study Center investigators based on instructions provided by WHO.

Experimental session. Experimental sessions were structured so as to minimize distractions and unintentional environmental influences. Participants were given a set of 60 laminated 9 cm by 4.5 cm white cards, each printed with the name of a mental disorder category phrased in a way that would be identifiable to both ICD-10 and DSM-IV users. See Table 1 for a list of the categories used in the study.[1] The cards were shuffled before beginning the session, so that the stack of cards was in random order when given to participants. The cards were numbered on the back to facilitate recording, as shown in Table 1, but the numbers were not visible to the participant.

Participants were then read the following instructions:

> In this study, we want to understand your views of the relationships among mental and behavioral disorders. Each of these cards contains the name of a mental disorder category. There are 60 cards. Categories have been labeled so that they are recognizable both to people who are familiar with ICD as well as the DSM. Slight differences between category names on these cards and the ones you might be familiar with are not intended to be meaningful. They are intended to correspond

[1] Because the experimental task would have been practically impossible had all possible disorders in the ICD and the DSM been used, we limited the number of disorders presented to participants. Before beginning the procedures for the current study, a pilot study identified exemplar disorders from the ICD and the DSM critical to understanding diagnostic taxonomies utilized by practitioners in clinical practice (see Roberts et al., 2012). This yielded the 60 disorder labels used in the study, listed in Table 1. The list is similar, but not identical, to the categories used in an earlier U.S. study of clinicians' natural taxonomies of mental disorders (Flanagan et al., 2008).

Table 1
Mental and Behavioral Disorders Category Labels Used in the Study

Card	Category	% Excluded	% Never used	% Remove
C1	Alzheimer's dementia	0.2%	5.0%	6.1%
C2	Vascular dementia	1.6%	10.9%	6.5%
C3	Amnestic disorder (organic)	2.1%	21.6%	7.5%
C4	Delirium	0.4%	7.3%	5.7%
C5	Mood disorder due to a general medical condition	0.2%	5.9%	2.2%
C6	Alcohol dependence	0.2%	1.8%	1.2%
C7	Opioid dependence	1.0%	13.9%	1.0%
C8	Cocaine dependence	0.4%	12.1%	0.8%
C9	Cannabinoid abuse	0.4%	9.5%	1.6%
C10	Abuse of volatile solvents (inhalants)	0.8%	18.0%	2.2%
C11	Tobacco (nicotine) dependence	0.4%	8.5%	5.0%
C12	Substance-induced psychotic disorder	0.6%	5.4%	0.2%
C13	Schizophrenia	0.0%	2.2%	0.0%
C14	Schizotypal disorder	1.4%	11.9%	1.8%
C15	Delusional disorder	0.2%	4.8%	0.6%
C16	Acute and transient (brief) psychotic disorder	0.4%	6.7%	0.6%
C17	Schizoaffective disorder	1.2%	5.2%	2.8%
C18	Bipolar I disorder	0.2%	2.0%	0.6%
C19	Bipolar II disorder	0.6%	1.4%	1.0%
C20	Depressive disorder (major)	0.2%	0.4%	0.0%
C21	Cyclothymia	2.7%	15.8%	5.0%
C22	Dysthymia	1.0%	3.6%	2.0%
C23	Panic disorder	0.0%	1.2%	0.2%
C24	Social phobia	0.0%	3.0%	0.6%
C25	Generalized anxiety disorder	0.0%	0.6%	0.8%
C26	Mixed anxiety and depressive disorder	0.8%	5.4%	4.2%
C27	Obsessive-compulsive disorder	0.0%	1.4%	0.2%
C28	Posttraumatic stress disorder	0.2%	3.4%	0.6%
C29	Adjustment disorders	2.5%	5.2%	5.7%
C30	Dissociative disorders	1.6%	9.9%	1.0%
C31	Conversion disorders	1.0%	7.1%	1.6%
C32	Somatization disorder	0.4%	3.8%	0.8%
C33	Hypochondriacal disorder	1.0%	7.5%	1.2%
C34	Persistent somatoform pain disorder	3.9%	17.0%	3.2%
C35	Body dysmorphic disorder	3.3%	18.6%	2.0%
C36	Anorexia nervosa	0.2%	8.1%	0.6%
C37	Primary (nonorganic) insomnia	1.4%	11.7%	8.9%
C38	Sexual dysfunction	1.2%	14.7%	10.3%
C39	Abuse of nondependence producing substances (e.g., steroids, hormones)	2.1%	28.3%	5.0%
C40	Paranoid personality disorder	0.0%	6.1%	2.0%
C41	Antisocial (dissocial) personality disorder	0.2%	5.0%	3.8%
C42	Borderline personality disorder	0.0%	3.4%	2.8%
C43	Dependent personality disorder	1.0%	9.5%	3.2%
C44	Pathological gambling	1.6%	26.9%	4.8%
C45	Intermittent explosive disorder	13.0%	38.6%	7.3%
C46	Paraphilias	7.0%	35.6%	9.1%

Table 1
Continued

Card	Category	% Excluded	% Never used	% Remove
C47	Gender identity disorder	3.1%	24.2%	19.4%
C48	Factitious disorder	10.3%	30.3%	4.6%
C49	Intellectual disability (mental retardation)	0.8%	4.4%	5.9%
C50	Specific developmental disorders of speech and language	2.1%	26.7%	6.7%
C51	Specific developmental disorders of scholastic skills	2.1%	25.2%	8.5%
C52	Autistic disorder	1.0%	13.5%	1.6%
C53	Asperger's syndrome	6.8%	23.8%	3.2%
C54	Attention deficit-hyperactivity (hyperkinetic) disorder	0.4%	4.8%	0.4%
C55	Conduct disorder	0.2%	7.3%	3.6%
C56	Oppositional defiant disorder	6.6%	22.6%	3.8%
C57	Childhood separation anxiety disorder	1.7%	21.0%	2.8%
C58	Reactive attachment disorder	11.0%	47.1%	6.3%
C59	Tic disorders	2.3%	11.9%	5.0%
C60	Nonorganic enuresis	3.1%	24.6%	8.5%

Note. "% Excluded" is the percentage of participants who *excluded* that category from sorting task based on lack of familiarity with the disorder. "% Never used" is the percentage who reported never having that category in clinical practice (i.e., who had never seen a patient with that diagnosis). "% Remove" is the percentage of participants who indicated that the category should not be included in a classification of mental and behavioral disorders.

> to the same diagnostic entities. I will ask you to arrange the cards to create groups of mental and behavioral disorders based on your own clinical experience of how similar they are, and how you approach the clinical management of these conditions. Your views may or may not match the structure of the ICD or DSM. We are interested in your views; this is not a test of your knowledge of existing systems. There are no right or wrong answers. Please sort the disorders on these cards into groups that you think are clinically relevant in the assessment and management of people with mental and behavioral disorders. There is no maximum or minimum number of groups, and each group can contain as few or as many disorders as you like. If you are unfamiliar with or lack basic knowledge about a particular disorder, you may set the card aside and move on to the next card.

The next step depended on the number of groups in the participant's first sorting. As described earlier, the goal was to obtain three levels of the participant's personal hierarchy of mental disorders at approximately the same level of detail that could reasonably be encompassed by a diagnostic classification system. Based on prior research (Flanagan et al., 2008), it was expected that the most common pattern would be for the participant to form more than four but fewer than ten groups in the first sorting. In that case, the experimenter read the following statement:

> Could you combine these groups into fewer, higher order groups that you think make sense clinically, without undoing the groups you have already formed? Some of these higher order groups might be the same as these original groups, if you believe it doesn't make sense to combine them with others into higher order groups.

Table 2
Demographic and Response-Related Variables by International Field Study Center

	Brazil	China	India	Japan	Mexico	Nigeria	Spain	U.S.A.	TOTAL
N Participants	60	62	61	73	67	60	74	60	517
Mean age (years)	40.6	38.8	39.0	41.6	43.7	42.4	42.8	49.1	42.3
Mean years of training	7.6	6.9	3.3	5.6	6.9	7.5	4.0	6.0	5.9
Mean years of experience	13.9	9.9	11.3	11.1	13.0	7.6	13.4	14.8	11.9
Patient contact per week (hours)	21.5	23.9	27.4	29.2	30.6	21.7	33.4	25.1	26.9
% psychiatrists	51.7%	100%	73.8%	100%	71.6%	68.3%	74.3%	40%	73.3%
% psychologists	46.7%	0%	21.3%	0%	28.4%	31.7%	18.9%	58.3%	24.8%
% used ICD-10 and DSM-IV	41.7%	11.3%	37.7%	60.3%	26.9%	38.3%	25.7%	6.7%	31.3%
% used ICD-10 only	13.3%	16.1%	54.1%	12.3%	3.0%	26.7%	16.2%	0.0%	17.6%
% used DSM-IV only	23.3%	16.1%	3.3%	19.2%	65.7%	30.0%	41.9%	78.3%	34.8%
% did not use ICD-10 or DSM-IV	21.7%	56.4%	4.9%	8.2%	4.5%	5.0%	16.2%	15.0%	16.2%
% participants who excluded categories	81.7%	64.5%	14.8%	20.5%	17.9%	16.7%	14.9%	25.0%	31.1%
Number of groups on first sorting	12.1	11.6	14.4	13.5	14.6	14.1	13.7	15.8	13.7
% participants who refused sorting	50.0%	43.5%	70.5%	6.8%	59.7%	56.7%	64.9%	31.6%	47.6%
Average maximum number of groups	13.2	14.8	16.4	17.6	14.7	15.9	13.3	17.3	15.8
Average minimum number of groups	6.0	6.9	6.6	6.5	8.6	6.6	7.4	7.4	6.9

Note. Average maximum number of groups and average minimum number of groups as shown above were calculated based *only* on those participants who did three sortings, as these participants provided the deepest hierarchical structure. Maximum corresponds to the number of groups in the participant's lowest order (most differentiated) sorting, and minimum corresponds to the number of groups in the participant's highest order (most aggregated) sorting. Maximum and minimum numbers of groups for participants who did two or one sorting fall between the maximum and minimum values for participants who did three sortings.

After the participant completed the aggregation or combining task to form higher order groups, the experimenter restored the original groups and asked the participant to perform a disaggregation or subdivision of the original groups:

> Could you subdivide these groups into more specific subgroups that you think would be clinically sensible? If it doesn't make sense to you to subdivide one of your original groups further, you can leave that group as it is.

Participants who generated more than 10 groups on the first sorting were asked to do two levels of aggregation or combining, that is, the aggregation step was completed twice, with the same instructions. According to the protocol, participants who generated four or fewer groups on the first sorting were to be asked to do two levels of disaggregation or subdividing.

Participant's groupings and rationale for each grouping at each level of aggregation were recorded. If, on the second or third sorting, participants indicated that it was not possible to combine or subdivide the groupings further, then the experimenter encouraged him or her to do so with a standard series of prompts, provided that the participant felt the resulting new groups would make sense clinically (for example, "What about this group? Do you feel it doesn't make sense to subdivide this group further?"). If the participant still indicated that she or he felt that existing groups could not be further combined or subdivided, then the experimenter proceeded to the third sorting (if the sorting refused was the second) or stopped the sorting task (if the sorting refused was the third).

After the sorting task had been completed, the experimenter reassembled all of the cards and asked the participant to go through them and indicate any disorders that he or she had never used in clinical practice (i.e., had never seen a patient with that diagnosis). The participant was then asked to go through the cards again and to indicate whether she or he felt that any of the 60 categories should not be included in a classification of mental and behavioral disorders and, if so, based on what rationale.

Finally, participants were asked to complete a brief questionnaire containing demographic, training, and practice information (available from the authors upon request). The total duration of the experimental session was between 60 and 90 minutes.

Analysis

The procedure described above produced a hierarchical arrangement of mental disorder categories for each participant. Relative distance among each pair of disorders was calculated for each participant based on the number of "steps" in the hierarchy before the disorders were joined in the same group. For example, if bipolar I disorder was grouped with bipolar II disorder in the lowest order sorting for a given participant (i.e., the most differentiated sorting with the greatest number of groups), then these two disorders would be given a distance of 1, meaning that they were in the same group at every level of the participant's hierarchy. If those two disorders were then joined with cyclothymia in the intermediate-level sorting, but not in the most differentiated sorting, then the distance between bipolar I disorder and cyclothymia would be 2. In turn, if all three of these disorders were grouped with depressive disorder (major) in the highest order sorting for a given participant (i.e., the most aggregated sorting with the smallest number of groups), then the distance between bipolar I and depressive disorder was 3. A maximum distance of 4 was used to represent disorders that were never grouped together.

This procedure produced a 60×60 disorder distance matrix for each participant, which were then averaged across all participants. This average matrix formed the basis of the cluster analysis described below. Average matrices for subgroups of participants were also constructed (i.e., by country, by profession, by which diagnostic manual participants most commonly used) and are discussed below.

Results

Demographic characteristics of the total sample and those recruited by each International Field Study Center are shown in Table 2. Of the 517 participants, 379 (73.3%) were psychiatrists, 128 (24.8%) were psychologists, and 10 (1.9%) were members of other disciplines (psychiatric nursing, counseling or social work). Consistent with the study objectives, the sample comprised experienced practicing clinicians, who had an average of 11.9 years of professional experience posttraining and spent an average of 26.9 hours per week providing direct mental health services to patients. There were some small but statistically significant differences across International Field Study Centers. U.S. clinicians were older than clinicians in all other countries, $F(7,509) = 6.83$, $p < .001$, and Nigerian clinicians had fewer years of experience than Brazilian, Mexican, Spanish, and U.S. clinicians (all ps $< .05$ after Tukey error correction).

As a part of the questionnaire, participants were asked to indicate on a 5-point scale, ranging from (*never*) to (*routinely*), how often they used the ICD-10 and the DSM-IV in their current clinical practice.[2] Across all participants, 162 (31.3%) reported using *both* the ICD-10 and the DSM-IV "routinely" or "often"; 91 (17.6%) used the ICD-10 "routinely" or "often" but not the DSM-IV; 180 (34.8%) used the DSM-IV "routinely" or "often" but not the ICD-10; and 84 (16.2%) reported that they used *neither* classification "routinely" or "often." As expected, there were large differences, $\chi^2(21) = 283.55$, $p < .001$, in whether participants used the ICD-10, the DSM-IV, both classifications, or neither classification by International Field Study Center, as shown in Table 2. Most of the participants who reported using neither classification were from the People's Republic of China, where the Chinese Classification of Mental Disorders (Chinese Society of Psychiatry, 2001)—an adaptation of the ICD-10—is the dominant classification system. The majority of participants who reported regularly using the ICD-10 but *not* the DSM-IV were from India, again reflecting practice patterns in that country. Most participants from the United States used the DSM-IV exclusively, and this pattern was common in Japan as well.

Across all participants, the average number of groups generated during the first sorting was 13.7 (standard deviation [*SD*] = 4.5, range = 5 – 28). Over three-quarters of the participants ($n = 392$; 75.8%) generated more than ten groups on their first sorting, with the remainder ($n = 125$; 24.2%) generating between five and ten groups. No participant generated four or fewer groupings of mental disorders on his or her first sorting. The second sorting was refused by 15 participants (2.9%), while 231 participants (44.7%) refused the third sorting. All participants who refused the second sorting were being asked to aggregate their initial categories at the time they refused the sorting, and this was true for 203 of the participants who refused the third sorting (87.9%). Only 28 (5.4%) of the participants who refused the third sorting were being asked to subdivide or disaggregate their groupings further at the time. All participants who included at least one sorting were included in data analyses.

Categories were excluded from the sorting task by 31.1% of participants because they were unfamiliar with them. The percentage of participants who excluded each category is shown in Table 1. When asked to indicate which categories they had never used in clinical practice (that is, had never seen a patient with that diagnosis), 83.8% of participants identified at least one category. At least one category was indicated as not belonging in a classification of mental and behavioral disorders by 60.4% of participants. The percentages of participants who indicated they had never used each category and who indicated that each category should not be included in a classification of mental disorders are shown in Table 2.

---

[2]The official system for diagnostic reporting in the US is the ICD-9 CM (Clinical Modification). Nine US participants reported "Routinely" or "Often" using the ICD-9; in all but one case these participants also "Routinely" or "Often" used the DSM-IV. The mental disorders classification in the ICD-9 CM is essentially the same as the DSM-IV, so use of the ICD-9 CM was not further considered in these analyses. Four US participants reported "Routinely" or "Often" using the ICD-10 in combination with the DSM-IV, as shown in Table 2.

Structure of the Classification

To examine the overall structure of mental disorders organization provided by clinicians, we submitted the distance matrices described above to cluster analysis using Ward's method. Ward's method was considered appropriate to the study questions because, compared to other cluster analytic methods, it tends to produce relatively homogenous groups that are similar in size (Aldenderfer & Blashfield, 1984). A rescaled cluster distance of 5 was adopted as a cutoff for determining what constituted a "group." Disorders that were joined at greater distances were not considered part of the same group. The resulting observed clusters are represented in Table 3. To provide a frame of reference, the clusters are listed in the order that most closely corresponds to ICD-10.

To examine the consistency of the obtained structure across participants in an intelligible and parsimonious way, we calculated a 60 × 60 matrix representing the frequency with which, at each participant's most differentiated sorting (i.e., the sorting with the highest number of groups), each disorder was grouped with every other disorder. Figure 1 shows the percentage of participants that placed each pair together in their most differentiated grouping (i.e., their sorting with the largest number of groups). The relative percentage of the sample that placed two disorders in the same group is also represented by shading in Figure 1, with darker shading indicating that a higher percentage of participants grouped those two disorders together.

The information presented in Figure 1 is directly relevant for understanding the classification structure generated by clinicians, but would be lost in a traditional depiction of cluster analysis results. Examining the frequency of disorder pairings in the manner depicted in Figure 1 has several advantages. First, it is a relatively face-valid, straightforward way of representing consensus among clinicians of how mental disorders should be organized in a classification. As can be seen in Figure 1, some disorders were grouped together by nearly the entire sample (darkest shading), other groupings were rare (lighter shading), and many disorders were never grouped together (blank cells).[3] Second, the information in Figure 1 provides a complete picture of the strength or cohesion of the clusters, and the covariance or overlap among disorders. This makes possible a direct examination of the distinctness or "fuzziness" of each disorder group and the pairings with disorders outside each cluster that account for this.

Stability of the Structure across Subgroups of Clinicians

To examine the stability of the observed structure as shown in Figure 1 across subgroups of participants, we calculated correlations for the averaged distance matrices by relevant subgroups. For example, to calculate a correlation among psychiatrists and psychologists, we averaged the matrices for members of each group, and then transformed them into single-column vectors. The correlation for the vectors was calculated in the usual way. It should be noted that these correlations are of averages, which will necessarily be higher than the correlations among individuals. Nonetheless, as a metric of reliability, the values can be interpreted.

Table 4 shows the correlations among averaged distance matrices across clinicians from different countries (International Field Study Centers). The consistency across global clinicians is remarkably high; all correlation values were above .90, indicating near linear redundancy among clinicians from the eight countries. Correlations for distance matrices by the five languages of administration were of similar magnitude, but are not presented here because language and country are obviously confounded. The diagnostic manuals that participants most commonly used did not affect the consensus matrix, with correlations among participant subgroups who used only the ICD-10, only the DSM-IV, both the ICD-10 and the DSM-IV, and neither classification were also uniformly above .90, as shown in Table 5. Finally, psychologists and psychiatrists were virtually identical in their consensus structures, $r = .97$.

[3]In Figure 1, cells corresponding to pairings made by fewer than .5% of participants (2 participants or less) were rounded to 0 and are shown as blank.

Table 3
Disorder Grouping Based on Cluster Analysis of Distance Matrices, With Cohesion Indices and Corresponding ICD-10 Categories

Observed clusters	Cohesion index	Corresponding ICD-10 categories
Neurocognitive and "organic" disorders	**0.47**	**F0**
Alzheimer's dementia	0.41	F00
Vascular dementia	0.41	F01
Amnestic disorder (organic)	0.41	F04
Delirium	0.45	F05
Mood disorder due to a general medical condition (organic)	0.67	F06.3
Substance-related disorders	**0.36**	**F1, part of F5**
Alcohol dependence	0.32	F10.2
Opioid dependence	0.32	F11.2
Cocaine dependence	0.32	F14.2
Cannabinoid abuse	0.33	F12.1
Abuse of volatile solvents (inhalants)	0.33	F18.1
Tobacco (nicotine) dependence	0.33	F17.2
Substance-induced psychotic disorder	0.61	F1x.5
Abuse of nondependence producing substances (e.g., steroids, hormones)	0.36	F55
Schizophrenia spectrum and other primary psychotic disorders	**0.47**	**F2**
Schizophrenia	0.41	F20
Schizotypal disorder	0.55	F21
Delusional disorder	0.43	F22
Acute and transient (brief) psychotic disorder	0.45	F23
Schizoaffective disorder	0.48	F25
Mood disorders	**0.38**	**F3**
Bipolar I disorder	0.37	F31
Bipolar II disorder	0.37	F31.8*
Depressive disorder (major)	0.39	F33
Cyclothymia	0.37	F34.0
Dysthymia	0.41	F34.1
Anxiety, obsessive-compulsive, and stress-related disorders	**0.60**	**Part of F4**
Panic disorder	0.51	F41.0
Social phobia	0.52	F40.1
Generalized anxiety disorder	0.51	F41.1
Mixed anxiety and depressive disorder	0.69	F41.2
Obsessive-compulsive disorder	0.62	F42
Posttraumatic stress disorder	0.59	F43.1
Adjustment disorders	0.74	F43.2
Dissociative and somatoform disorders	**0.55**	**Parts of F4 & F6**
Dissociative disorders	0.59	F44
Conversion disorders	0.51	F44
Somatization disorder	0.49	F45.0
Hypochondriacal disorder	0.54	F45.2
Persistent somatoform pain disorder	0.51	F45.4
Factitious disorder	0.68	F68.1
Other bodily disorders	**0.86**	**Parts of F4 & F5**
Body dysmorphic disorder	0.86	F45.2*
Anorexia nervosa	0.81	F50.0
Primary (nonorganic) insomnia	0.90	F51.0
Personality disorders	**0.41**	**Part of F6**
Paranoid personality disorder	0.44	F60.0

Table 3
Continued

Observed clusters	Cohesion index	Corresponding ICD-10 categories
Antisocial (dissocial) personality disorder	0.40	F60.2
Borderline personality disorder	0.39	F60.31
Dependent personality disorder	0.40	F60.7
Disorders of impulse control	**0.78**	**Part of F6**
Pathological gambling	0.78	F63.0
Intermittent explosive disorder	0.78	F63.8*
Sexual disorders	**0.54**	**Parts of F5 & F6**
Sexual dysfunction	0.56	F52
Paraphilias	0.51	F65
Gender identity disorder	0.55	F64
Neurodevelopmental disorders	**0.47**	**F7, F8, part of F9**
Intellectual disability (mental retardation)	0.55	F7
Specific developmental disorders of speech and language	0.43	F80
Specific developmental disorders of scholastic skills	0.44	F81
Autistic disorder	0.45	F84.0
Asperger's syndrome	0.46	F84.5
Attention deficit-hyperactivity (hyperkinetic) disorder	0.52	F90
Other childhood disorders	**0.69**	**Most of F9**
Conduct disorder	0.68	F91
Oppositional defiant disorder	0.65	F91.3
Childhood separation anxiety disorder	0.63	F93.0
Reactive attachment disorder	0.63	F94.1
Tic disorders	0.80	F95
Nonorganic enuresis	0.73	F98.0

Note. Cohesion index is calculated based on average within group distances divided by average distances from categories outside group. Therefore, higher values for cohesion index (closer to 1) indicate *less* cohesion. Disorder categories for which corresponding ICD-10 categories indicated are asterisked (*) above do not appear in ICD-10 as separate disorders but are inclusion terms under the category number provided.

Cohesion Across Disorder Clusters

The differences between the observed consistency of some disorder groupings as shown in Figure 1 and the extremely high correlations shown in Tables 5 and 6 suggest that although the *average* matrix across subgroups is essentially identical, there is substantially more *individual* variability and overlap in some clusters than in others. For example, Figure 1 shows that the observed cluster of substance-related disorders is highly cohesive, with more than 85% of participants grouping all of these disorders together in their most differentiated sorting, with the exception of substance-induced psychotic disorder, which was grouped with schizophrenia spectrum and other primary psychotic disorders by approximately half of participants. On the other hand, the observed cluster of anxiety, obsessive-compulsive, and stress-related disorders is much less cohesive. Although this cluster contains a highly cohesive subcluster comprising panic disorder, social phobia, and generalized anxiety disorder, some disorders (e.g., adjustment disorder) were maintained in the cluster by only a minority of participants. This observed individual variability may have important implications for classification, as it could affect the consistency with which groupings of disorders are conceptualized and implemented across clinicians.

Table 4
Correlations Among Average Clinician Distance Matrices by International Field Study Center

	Brazil	China	India	Japan	Mexico	Nigeria	Spain
China	.911						
India	.943	.962					
Japan	.921	.960	.956				
Mexico	.963	.911	.959	.920			
Nigeria	.943	.951	.977	.946	.968		
Spain	.945	.943	.968	.951	.962	.966	
USA	.963	.901	.939	.917	.959	.944	.939

Table 5
Correlations Among Average Clinician Distance Matrices by Diagnostic System Currently Used

	Use ICD only	Use DSM only	Use both
Use DSM only	.980		
Use both	.991	.987	
Use neither	.982	.966	.979

While cluster analysis provides an overall picture of how participating clinicians considered the mental disorders should be grouped together, it does not describe the relative cohesion of the groups of disorders that were created. In other words, some groups might be very tightly knit, having low in-group distances relative to larger out-group distances. Others might be loose, having in-group distance values similar to out-group distance values. At the extreme end, the members of a group might be no closer to one another than they are to disorders outside the group. To examine this effect, we calculated a "cohesion index" for each *group* of disorders by dividing the average distance among disorders within the group by the average distance from disorders not in the group. The cohesion index was calculated for individual *disorders* by dividing the average difference between that disorder and other disorders within the group by its average distance from disorders outside the group. Low values on the cohesion index (closer to 0) indicate strong group cohesion. High values (near to 1) indicate poor group cohesion. Cohesion index values for each observed grouping and for each individual disorder are shown in Table 3. If a particular grouping is not cohesive based on a relatively high cohesion index value shown in Table 3, or a particular disorder is not cohesive with the other disorders in the observed grouping, then it is possible to see which other disorders the grouping overlaps by examining Figure 1.

As noted, substance-related disorders was the most cohesive groping (cohesion index = .36). The mood disorders grouping was also quite cohesive (.38), as was personality disorders (.41), with little variability among their members. The grouping called schizophrenia spectrum and other primary psychotic disorders was also fairly cohesive (.47), with schizotypal disorder and schizoaffective disorder less tightly attached to the group than the other members. The grouping of anxiety, obsessive-compulsive, and stress-related disorders and the grouping of dissociative and somatoform disorders had only moderate cohesion (.60 and .55, respectively), primarily due to their substantial overlap with each other and with mood disorders.

Not surprisingly, the group labeled other bodily disorders evidenced the poorest internal cohesion. A cohesion index of .86 indicates that the disorders within the group were only slightly closer to one another than they were to disorders outside the group, as can also be seen in Figure 1. This group of disorders combines several distinct areas in current diagnostic schemes, and no disorders that would be more commonly paired with members of this group (e.g., bulimia nervosa with anorexia nervosa) were included in the study, making it impossible for separate groups to emerge. Thus, this grouping is likely a methodological artefact of including a limited number of stimuli. Similarly, pathological gambling and intermittent explosive disorder

Disorder Cluster Group	Diagnostic Category	Alzheimer's dementia	Vascular dementia	Amnestic disorder (organic)	Delirium	Mood disorder due to a medical condition	Alcohol dependence	Opioid dependence	Cocaine dependence	Cannabinoid abuse	Abuse of volatile solvents (inhalants)	Tobacco (nicotine) dependence	Substance-induced psychotic disorder	Abuse of non-dependence producing substances	Schizophrenia	Schizotypal disorder	Delusional disorder	Acute and transient (brief) psychotic disorder	Schizoaffective disorder	Bipolar I disorder	Bipolar II disorder	Depressive disorder (major)	Cyclothymia	Dysthymia
		Neurocognitive and 'organic'					Substance-related								Schizophrenia and psychotic						Mood					
Neurocognitive and 'organic'	Alzheimer's dementia	--	98	83	70	39		1		1	1		4	1			1		1			1		
	Vascular dementia	98	--	82	71	39		1		1	1		4	1			1							
	Amnestic disorder (organic)	83	82	--	78	46	1	1	1	1	1	1	6	1			1	1				1		1
	Delirium	70	71	78	--	43	1	1	1	1	1	1	10	1	4	3	4	5	3	1	1	1	1	
	Mood disorder due to a medical condition	39	39	46	43	--	1	1	1	1	1	1	9	1	1	1	1	1	4	33	33	37	32	34
Substance-related	Alcohol dependence			1	1	1	--	99	99	95	95	97	54	87										
	Opioid dependence	1	1	1	1	1	99	--	99	95	94	97	53	86										
	Cocaine dependence			1	1	1	99	99	--	95	95	98	54	87										
	Cannabinoid abuse	1	1	1	1	1	95	95	95	--	98	94	53	90										
	Abuse of volatile solvents (inhalants)	1	1	1	1	1	95	94	95	98	--	93	54	90										
	Tobacco (nicotine) dependence			1	1	1	97	97	98	94	93	--	53	86										
	Substance-induced psychotic disorder	4	4	6	10	9	54	53	54	53	54	53	--	49	29	20	28	30	22	2	2	1	1	
	Abuse of non-dependence producing substances	1	1	1	1	1	87	86	87	90	90	86	49	--										
Schizophrenia and psychotic	Schizophrenia				4	1							29		--	66	83	79	77	7	6	3	3	1
	Schizotypal disorder				3	1							20		66	--	60	56	59	6	5	3	4	2
	Delusional disorder	1	1	1	4	1							28		83	60	--	78	67	5	5	3	3	1
	Acute and transient (brief) psychotic disorder			1	5	1							30		79	56	78	--	64	5	5	3	4	1
	Schizoaffective disorder	1			3	4							22		77	59	67	64	--	18	17	11	13	9
Mood	Bipolar I disorder				1	33							2		7	6	5	5	18	--	97	74	83	66
	Bipolar II disorder				1	33							2		6	5	5	5	17	97	--	75	83	66
	Depressive disorder (major)	1		1	1	37							1		3	3	3	3	11	74	75	--	74	85
	Cyclothymia				1	32							1		3	4	3	4	13	83	83	74	--	79
	Dysthymia				1	34									1	2	1	1	9	66	66	85	79	--
Anxiety, obsessive-compulsive, and stress-related	Panic disorder			1	1	4									1	1	1	2	1	5	5	9	7	12
	Social phobia				1	3									1	1	2	1	1	4	4	7	6	10
	Generalized anxiety disorder					5									1		1	1	1	6	7	12	9	15
	Mixed anxiety and depressive disorder	1		1		23									1	1	1	1	6	38	38	50	44	53
	Obsessive-compulsive disorder				1	3	1	1	1			1	1	1	2	4	3	2	2	5	5	6	6	9
	Posttraumatic stress disorder	1	1	1	1	4											1	3	1	5	5	7	7	10
	Adjustment disorders					7								1		1		3	1	10	10	14	13	17
Dissociative and somatoform	Dissociative disorders	1	1	2	2	1							1		6	6	6	6	4	2	2	2	4	5
	Conversion disorders				1	1									1	4	3	2	1	2	1	2	3	5
	Somatization disorder			1	1	3								1	1	1	2	1	1	2	2	3	3	5
	Hypochondriacal disorder			1		2	1	1	1		1	1	2	1	2	4	4	2	2	3	3	4	4	6
	Persistent somatoform pain disorder			1	1	2									1	1	2	1	1	2	2	3	3	5
	Factitious disorder			1	1	2							1	1	1	3	2	2	1		1	1	2	3
Other bodily	Body dysmorphic disorder			1	1	1							3	2	9	8	13	10	7	2	2	2	3	3
	Anorexia nervosa				1	1	1	1	1	1	1	1	1	3	1	2	2	2	1	2	2	2	2	3
	Primary (nonorganic) insomnia	1	1	2	2	4				1	1			1		1		1		3	3	5	4	6
Personality	Paranoid personality disorder				2	1							7		17	45	20	15	17	3	3	1	3	2
	Antisocial (dissocial) personality disorder				1	1							1		1	27	2	1	3	1	1	1	2	2
	Borderline personality disorder				1	1							2	1	2	30	4	3	5	3	4	2	4	4
	Dependent personality disorder				1	1	2	2	2	2	2	2	2	3	1	27	2	2	3	1	1	2	3	4
Impulse control	Pathological gambling				1	1	34	33	34	33	33	34	14	31		3	1	1	1	1	1	1	1	1
	Intermittent explosive disorder	1	1	2	1	3	1	1	1	1	2	1	1	2	2	11	3	5	2	4	5	5	4	4
Sexual	Sexual dysfunction			1	1	1				1	1					1	1	1		1	1	2	2	3
	Paraphilias			1	1		2	1	1	2	1	1		1	1	2	2	1	1			1	2	1
	Gender identity disorder													1		3	1	1					1	1
Neurodevelopmental	Intellectual disability (mental retardation)	8	8	8	7	5							1											
	Specific devel. disorders of speech & language	1	1	1	1	1																		
	Specific devel. disorders of scholastic skills	1	1	1		1																		
	Autistic disorder	2	2	2	2	2									2	2	2	2	2					
	Asperger's syndrome	5	5	5	5	3									1	2	1	1	1	1	1			
	ADHD (hyperkinetic disorder)	1	1	1	1	1											1	1						1
Other childhood	Conduct disorder					1							1		1	5	1	1	1	1	1	1		1
	Oppositional defiant disorder															4	1	1	1	1	1		1	1
	Childhood separation anxiety disorder					1										1		1		2	2	3	2	3
	Reactive attachment disorder															1		1			1	1	1	2
	Tic disorders	5	5	6	6	6	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	3	2	2
	Nonorganic enuresis				1	1												1				1	1	1

Cell shading: 10 20 30 40 50 60 70 80 90 100

Percentage of participants: 0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

Figure 1. Percentage of participants (n = 517) grouping each disorder pairing in their most differentiated grouping.

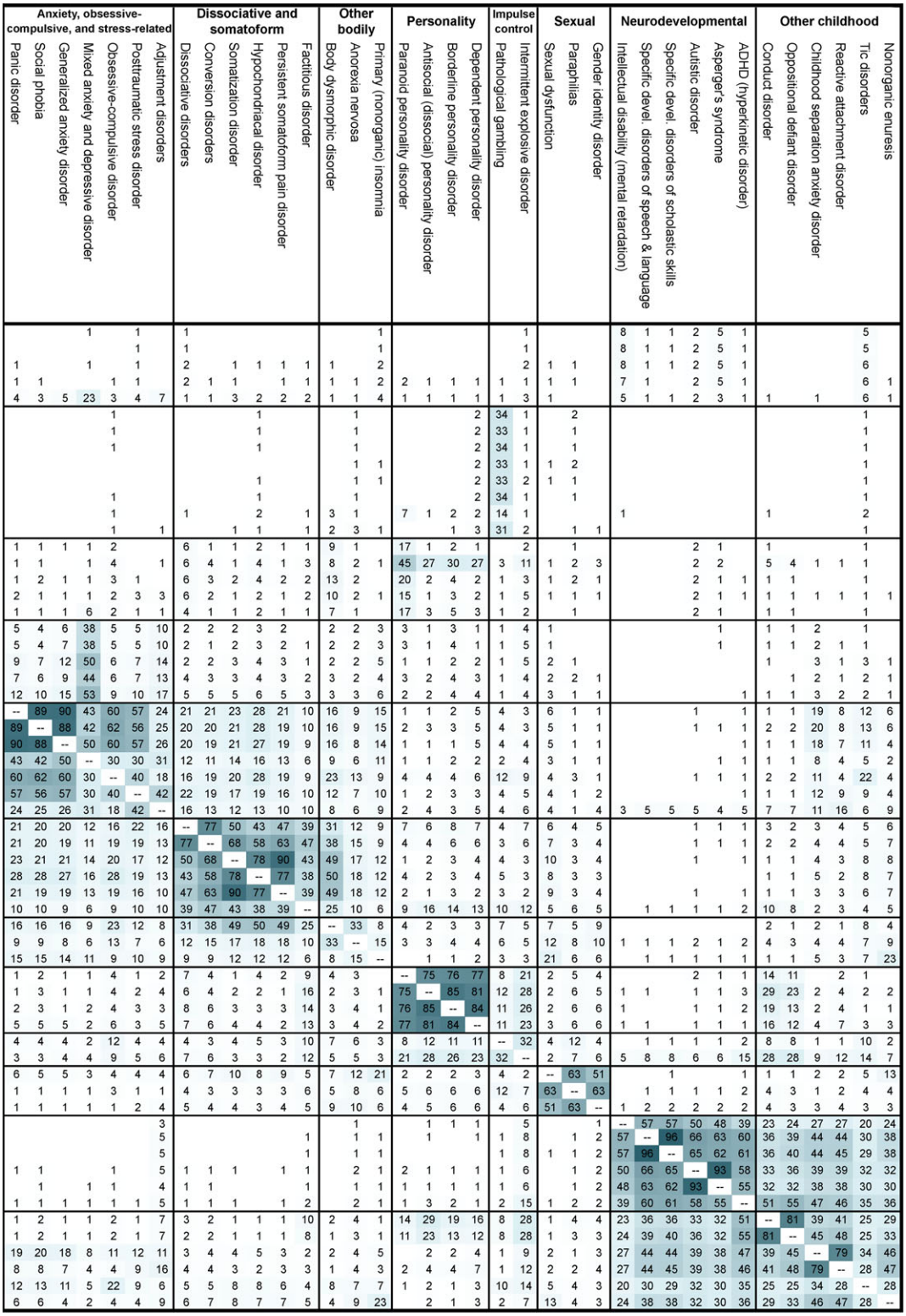

Figure 1. Continued

were classificatory misfits, in that they were only slightly closer to each other than they were to any other disorder (.78). In fact, Figure 1 shows that participants more commonly grouped pathological gambling with substance-related disorders. The grouping of other childhood disorders was more cohesive than these two, but less cohesive than any of the other groupings (.69), showing overlap not only with neurodevelopmental disorders but also to a lesser degree with a diffuse set of conditions in other groupings (see Figure 1).

Discussion

This study revealed a "natural taxonomy" of mental disorders held by global mental health professionals that was strikingly consistent, with correlation coefficients higher than .90 across countries (Table 4), languages, classification system used (Table 5), and professional discipline. In the context of considerable discussion about the lack of reliability among clinicians (e.g., Garb, 2005; Regier et al., 2013), these results are compelling. Clinicians interact with people with mental and behavioral disorders on a daily basis and form implicit (or sometimes explicit) views of the relationships among disorders (Egli et al., 2006; Flanagan et al., 2008; Roberts et al., 2012). At the same time, clinicians' perspectives are obviously shaped by training and theoretical and practical knowledge about the nature of psychopathology, which is increasingly shared throughout the world. The ways in which this clinician-generated classification structure deviates from current classification systems are neither random nor idiosyncratic, but they are strongly shared across countries, languages, and the professional disciplines of psychiatry and psychology. Although the classification structure generated by clinicians through this task has clear similarities to both ICD-10 and DSM-IV, it is distinct from either of these systems (see Table 3 and Figure 1). Clinicians were not simply parroting back the classification systems they had been taught.

A major purpose of this study was to consider how clinicians' organizations of mental disorders might be used to inform revisions of ICD, and several characteristics of clinicians' groupings as compared to ICD-10 and proposals for ICD-11 seem particularly relevant. Global clinicians' groupings of neurocognitive and "organic" disorders, substance-related disorders, schizophrenia spectrum and other primary psychotic disorders, and mood disorders corresponded quite closely to disorder groupings in ICD-10 (F0, F1, F2, and F3), which will essentially be retained in ICD-11. Clinicians included schizotypal disorder in the schizophrenia spectrum and other primary psychotic disorders grouping, as in ICD-10 and as proposed for ICD-11, and not with the personality disorders as in DSM-IV. The groupings of neurodevelopmental disorders and other childhood disorders correspond in essence to the F8 and F9 grouping in ICD-10, with the category corresponding to mental retardation[4] (F7) in ICD-10 absorbed into clinicians' neurodevelopmental disorders grouping. The neurodevelopmental disorders grouping generated by clinicians is nearly identical to the neurodevelopmental disorders grouping proposed for ICD-11. Interestingly, attention deficit-hyperactivity disorder (ADHD) was included in this grouping. One difference is that the current ICD-11 proposal also includes tic disorders in the neurodevelopmental grouping. While clinicians placed tic disorder in the other childhood disorders grouping, it was paired with the categories in neurodevelopmental disorders with approximately the same frequency (see Figure 1). The recommendation has been made *not* to include a grouping of other childhood disorders in ICD-11, but rather to group these with the corresponding "adult" disorders to provide a more developmental perspective (Rutter, 2011).

Three of the major groupings in ICD-10 are very broad and quite heterogeneous. These include: F4, neurotic, stress-related, and somatoform Disorders; F5, behavioural syndromes associated with physiological disturbances and physical factors; and F6, disorders of adult personality and behaviour. In the present study, clinicians divided the categories in these groupings into a larger number of more narrowly defined groups. This finding is consistent with proposals for ICD-11. Global clinicians grouped mixed anxiety-depressive disorder with anxiety disorders, as in ICD-10, although they placed it with mood disorders with approximately the same

[4]In the current study, the category was labeled "Intellectual disability (mental retardation)." The proposed category name for ICD-11 is "Disorders of intellectual development."

frequency (see Figure 1), as has been proposed for ICD-11. Only 4.2% of clinicians said that mixed anxiety-depressive disorder should not be included in a classification of mental and behavioral disorders (see Table 1).

There is little evidence based on these data that clinicians conceptualize a grouping of obsessive–compulsive and related disorders, as has been proposed for ICD-11. Of possible candidates for such a grouping, obsessive-compulsive disorder was paired with hypochondriacal disorder by 27.9% of participants, with body dysmorphic disorder by 22.6% and tic disorders by 22.0%, while having much higher co-occurrence with anxiety disorders.

Any classification system of mental disorders serves the dual roles of representing the science of psychopathology as accurately as possible given the current state of knowledge (i.e., validity) while also maintaining utility for the clinicians who implement it on a daily basis. A scientifically valid classification that is unusable in real-world health care settings will not lead to better implementation of standardized classification in clinical settings, nor will it lead to improved quality of health encounter data for the variety of purposes for which it is used. The current study represents an attempt to systematize clinicians' collective experience with psychopathological conditions and their relationship to one another *as one possible source* of information that will be useful in making decisions about how to structure the ICD-11 classification.

An obvious question arising from these results concerns the relationship between clinicians' implicit classification and patient-level phenomena. Clinicians' implicit classifications may represent an accurate reflection of some patient variables, and a distortion or misinterpretation of others. The consistency of the implicit classification found in this study across clinicians suggests that it is at least accurate enough to be reliably reproduced across countries, languages, and professions. It implies that disorders are real and accurately perceived, provided that clinicians have the necessary information. This is likely to vary across individual patients and individual clinicians based on clinician expertise, the adequacy of the assessment or observation of relevant diagnostic symptoms, and the extent to which relevant diagnostic symptoms for particular disorders are observable.

We are not suggesting that the conceptualizations of clinicians should override other forms of evidence. In those cases in which compelling and dispositive data are available, the classification should be structured according to the scientific evidence. But, even then, understanding clinicians' cognitive structures for mental disorders classification can be important because it will aid substantially in identifying those places of substantial divergence that should be a particular target of educational efforts at the time the new classification is implemented. In the absence of evidence that contradicts clinicians' natural taxonomies, however, the extent to which alternative organizations facilitate the use of the classification in clinical settings should be among the factors that are weighed in decisions about the final structure.

Understanding clinicians' working taxonomies is of practical importance because these are the same individuals who will be asked to apply a standardized system for identifying and treating mental disorders, often under extreme pressure and with very little time for detailed diagnostic interviewing. As noted above, many of the changes already being countenanced for the ICD-11 are consistent with the clinician-generated structure. It is reasonable to assume—and a testable hypothesis—that the more compatible the organization of the classification is with clinicians' implicit structures of mental and behavioral disorders categories, the easier and more intuitive it will be for them to use the system.

Some implications may also be drawn from this study regarding the general characteristics of a classification that clinicians might prefer. Clinicians in this study preferred a flatter organizational structure for grouping mental disorders, with a relatively large number of groups. The average number of groups clinicians generated on the first sorting trial, which could be seen as reflecting their most spontaneous or natural organization, was nearly 14. Moreover, clinicians did not create deeply hierarchical structures even within the limits set by the study methodology. Nearly half the participants overall (47.6%) did not take full advantage of the hierarchical structure available to them and refused at least one sorting, despite prompts and encouragement by the experimenter. Participants were more resistant to hierarchical aggregation than disaggregation, with nearly all of the refused sortings occurring when participants were being asked to combine groups into larger, higher order groups. Of participants who did use the fully available three-level

hierarchy, the number of groups they created at the highest order (most aggregated) level was about seven (see Table 1), while their most detailed, lowest order level of organization contained approximately 16 groups. There is certainly no evidence that clinicians conceptualize the sort of radical collapsing of "emotional" and "externalizing" disorders, proposed by Andrews and colleagues (Andrews et al., 2009; Goldberg, Krueger, Andrews, & Hobbs, 2009; Krueger & South, 2009).

The wide range of experienced clinicians participating in this study—working in different contexts in different countries and participating in different languages—is one of its major strengths. Participants had an average of nearly 6 years of training and 12 years of professional experience posttraining. Participants were practitioners, with an average of almost 27 hours of patient contact per week. Participants were not randomly selected for participation, but the 517 clinicians who participated in this study can be conceptualized as an alternative (and large) international panel of experts, supplementing the nonrandom panels collected by WHO and the American Psychiatric Association to develop their respective classifications. The emphasis in this study on the direct participation of clinicians from non-Anglophone countries outside North America and Western Europe provides a basis for examining the cross-cultural applicability of disorder concepts that have been considered "Western."

There are also some limitations that should be considered in interpreting the study's results. The most important one relates to the relatively limited set of disorder stimuli, which likely limited the number of groups that emerged in the clinician sortings. It seems likely that had some additional disorders been included (e.g., bulimia nervosa), additional groupings might have emerged. In addition, the placement of some disorders may have been related to clinicians' lack of knowledge about them. For example, 47.1% of participants reported never having seen someone with reactive attachment disorder in clinical practice, and 38.6% had never encountered a patient with intermittent explosive disorder (which is not a separate category in the ICD-10), though only 11% and 13%, respectively, admitted that they were not familiar enough with these categories to be able to sort them.

Further, it is not entirely possible to separate participant's "natural" taxonomies of mental disorders from their own training in particular systems of classification (ICD or DSM). These formal classification systems likely act as a baseline for clinicians' "natural" classifications. The similarity between clinicians' classifications and those of diagnostic manuals is doubtless in part due to their training and not just because this is what they have learned through their own experience. On the other hand, mental disorders classifications began as attempts to systematize clinicians' observations and therefore likely have some degree of naturalistic validity such that when clinicians agree with the system, it is in part a reflection of this validity and not just a result of rote training. In the present study, what is perhaps most important features of the data are the points of departure observed in clinicians' taxonomies from existing classification systems, and the consistency of these deviations across countries, languages, and professions, regardless of which classification system they used in daily practice.

Conclusions

The most striking feature of this study of the natural taxonomy of mental disorders among 517 mental health professionals from around the world was the consistency and robustness of the clinician-generated model. Despite all the myriad ways in which individual clinicians might differ, they constructed a classification of mental disorders that was virtually identical across psychiatrists and psychologists in Brazil, China, India, Japan, Mexico, Nigeria, Spain, and the United States, whether they were asked in Chinese, English, Japanese, Portuguese or Spanish. This commonality suggests a communal understanding of mental disorders among clinicians that appears to transcend cultural or professional differences. We are not suggesting that there are no cultural differences in the expression of psychopathology or that there are no meaningful distinctions among professions. Rather, we wish to highlight that the common organization of mental disorders obtained in this study may be relevant to the development of classifications of mental and behavioral disorders because it can be used to improve the interface between health information systems and clinical practice, so that (a) practice is more usefully guided by standard

concepts and definitions, (b) data generated in health encounters reflect clinical experience more accurately, and (c) aggregated health encounter data are more valid and reliable for decision making and policy setting by health systems and governments. Attention to these issues will help to make the ICD-11 a more effective tool for identifying people who need mental health treatment and reducing the burden of mental disorders throughout the world.

In particular, research is needed to address the enormous unmet mental health needs of low- and middle-income countries (e.g., WHO, 2012). Science, in its quest to accomplish valid generalizations about nature, is inherently global. Researchers from all parts of the world should, desirably, contribute to new knowledge about mental health and mental illness. Mental health research from low- and middle-income countries is needed for acceptance of classification systems in these countries. A steady stream of information about mental health issues in these countries will contribute to a greater international and multicultural understanding of mental disorders and of mental health.

References

Aldenderfer, M. S., & Blashfield, R. K. (1984). Cluster Analysis. Newbury Park, CA: Sage.

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.

Berlin, B. (1992). Ethnobiological classification. Princeton, NJ: Princeton University Press.

Blashfield, R. K., Keeley, J., & Burgess, D. (2009). Classification.In P. Blaney & T. Millon (Eds.), Oxford textbook of psychopathology (2nd ed., pp. 35–57). New York: Oxford University Press.

Cantor, N., Smith E. E., French, R., & Mezzich, J. (1980). Psychiatric diagnosis as prototype categorization. Journal of Abnormal Psychology, 89, 181–193.

Chinese Society of Psychiatry. (2001), Chinese classification of mental disorders (3rd ed., CCMD-3). Shandong, China: Shandong Publishing House of Science and Technology.

Egli, S., Schlatter, K., Streule, R., & Läge, D. (2006). A structure-based expert model of the ICD-10 mental disorders. Psychopathology, 39, 1–9.

Egli, S., Streule, R., & Läge, D. (2008). The structure-based expert model of the mental disorders: A validation study. Psychopathology, 41, 286–293.

First, M. B. (2009). Reorganizing the diagnostic groupings in DSM-V and ICD-11: A cost/benefit analysis. Psychological Medicine, 39, 2091–2097.

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. H., & Blashfield, R. K. (2006). Do clinicians see Axis I and Axis II as different kinds of disorders? Comprehensive Psychiatry, 47, 496–502.

Flanagan, E. H., & Blashfield, R. K. (2007). Clinicians' folk taxonomies of mental disorders. Journal of Philosophy, Psychiatry, and Psychology, 14, 249–269.

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

Flanagan, E. H., Keeley, J., & Blashfield, R. K. (2012). Do clinicians conceptualize DSM-IV disorders hierarchically? Journal of Clinical Psychology, 68, 620–630.

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

Goldberg, D. P., Krueger, R. F., Andrews, G., & Hobbs, M. J. (2009). Emotional disorders: Cluster 4 of the proposed meta-structure for DSM-V and ICD-11. Psychological Medicine, 39, 2043–2059.

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

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, 10, 86–92.

Jablensky, A. (2009). A meta-commentary on the proposal for a meta-structure for DSM-V and ICD-11. Psychological Medicine, 39, 2099–2103.

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

Krueger, R. F., & South, S. C. (2009). Externalizing disorders: Cluster 5 of the proposed meta-structure for DSM-V and ICD-11. Psychological Medicine, 29, 2061–2070.

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.

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

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.

Regier, D. A., Narrow, W. E., Clarke, D. E., Kraemer, H. C., Kuramoto, S. J., Kuhl, E. A., & Kupfer, D. J. (2013). DSM-5 field trials in the United States and Canada, part II: Test-retest reliability of selected categorical diagnoses. American Journal of Psychiatry, 170, 59–70.

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.

Rutter, M. (2011). Child psychiatric diagnosis and classification: Concepts, findings, challenges and potential. Journal of Child Psychology and Psychiatry, 52, 647–660.

Wittchen, H.-U., Beesdo, K., & Gloster, A. T. (2009). A new meta-structure of mental disorders: A helpful step into the future or a harmful step back to the past? Psychological Medicine, 39, 2083–2089.

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

World Health Organization. (2012). World Health Assembly Resolution 65.4: Global burden of mental disorders and the need for a comprehensive, coordinated response from health and social sectors at the country level. Geneva, Switzerland: Author.