

A global clinicians' map of mental disorders to improve ICD-11: Analysing meta-structure to enhance clinical utility

MICHAEL C. ROBERTS¹, GEOFFREY M. REED², MARÍA ELENA MEDINA-MORA³, JARED W. KEELEY⁴, PRATAP SHARAN⁵, DAVID K. JOHNSON⁶, JAIR DE JESUS MARI⁷, JOSÉ LUIS AYUSO-MATEOS⁸, OYE GUREJE⁹, ZEPING XIAO¹⁰, TOSHIMASA MARUTA¹¹, BRIGITTE KHOURY¹², REBECA ROBLES³ & SHEKHAR SAXENA²

¹Clinical Child Psychology Program, University of Kansas, Lawrence, Kansas, USA, ²Department of Mental Health and Substance Abuse, World Health Organization, Geneva, Switzerland, ³National Institute of Psychiatry Ramón de la Fuente, Mexico, ⁴Department of Psychology, Mississippi State University, Starkville, Mississippi, USA, ⁵Department of Psychiatry, All India Institute of Medical Sciences, New Delhi, India, ⁶Department of Psychology and Gerontology Center, University of Kansas, Lawrence, Kansas, USA, ⁷Department of Psychiatry, Universidade Federal de São Paulo, Brazil, ⁸Department of Psychiatry, Universidad Autónoma de Madrid, Instituto de Investigación Sanitaria Princesa, CIBERSAM, Madrid, Spain, ⁹Department of Psychiatry, University of Ibadan, Ibadan, Nigeria, ¹⁰Shanghai Mental Health Centre, Shanghai Jiaotong University School of Medicine, Shanghai, China, ¹¹Department of Psychiatry, Tokyo Medical University, Tokyo, Japan, and ¹²Department of Psychiatry, American University of Beirut, Lebanon

Abstract

Enhancing clinical utility is an emphasis of the World Health Organization's development of the mental and behavioural disorders chapter of the next *International Classification of Diseases* (ICD-11). Understanding how clinicians conceptualize the structure of mental disorders can enable a more clinically intuitive classification architecture that will help professionals find the categories they need more efficiently. This study examined clinicians' conceptualizations of the relationships among mental disorders and the dimensions they use in making these judgements. Psychiatrists and psychologists from 64 countries (n = 1,371), participating in English or Spanish, rated the similarity of mental and behavioural disorders presented as paired comparisons. Data were analysed by multidimensional scaling procedures (INDSCAL) and by analyses of consistency. Participants used three distinctive dimensions to evaluate the similarity among disorders: internalizing versus externalizing, developmental versus adult onset, and functional versus organic. Clinicians' conceptual map of mental disorders was rational and highly stable across profession, language, and country income level. The proposed ICD-11 structure is a moderately better fit with clinicians' conceptual model than either ICD-10 or DSM-IV. Clinician judgements can be used to improve clinical utility of the ICD-11 without sacrificing validity based on a scientific approach to enhancing a logically organized classification meta-structure.

Introduction

Current classification systems for mental and behavioural disorders are widely acknowledged to suffer from major problems of clinical utility (First, 2010; Hyman, 2010; International Advisory Group, 2011; Kendell & Jablensky, 2003). Improving the clinical utility of the ICD-11 is a major goal of the World Health Organization (WHO). This focus is directly related to WHO's priority as a global public health agency to assist its 194 member countries to improve the rate and effectiveness of treatment and to reduce the disease burden associated with mental

and behavioural disorders. Clinical utility is connected to the way that the classification is organized. The groupings of disorders and the classification's hierarchical structure should be logical and empirically supported (i.e. valid), but should also be intuitive and accessible for clinicians around the world (First, 2010; Flanagan & Blashfield, 2010; Reed, 2010).

Global healthcare systems are overburdened, and clinicians are under too much time pressure to engage in detailed diagnostic interviewing. If a classification is difficult and cumbersome to implement and does

not provide information that is of immediate value to clinicians in real-world healthcare settings it has no hope of being implemented carefully and accurately at the encounter level. Data generated in the context of health encounters are used by governments and healthcare systems for many purposes, including morbidity reporting, allocation of healthcare resources, the development of clinical and public health programmes, and as a basis for health policy, such as in eligibility determination and reimbursement. A diagnostic system that is not useful at the encounter level cannot generate information that will be valid for these purposes. It is unclear that attempts to organize a classification of mental and behavioural disorders based on the consensus of a group of experts serving on a classification committee have led to classification structures that best meet the need for clinical utility at the encounter level and best meet the public health objectives of WHO. The complexity of the definitions and diagnostic guidelines for mental disorders and the opacity and inconsistency of the rationale for how diagnostic categories are organized in existing classifications are likely major elements in the poor clinical utility of existing classifications (Andrews et al., 2008; First, 2010; Reed, 2010).

The organization of disorders must be useful to practitioners all over the world, ranging from highly trained mental health professionals to primary care service providers with little specific training in mental and behavioural disorders. A more clinically useful structure would be one that helps clinicians to find, as quickly and efficiently as possible, the diagnostic category that best applies to the person in front of them and is most likely to provide useful information about treatment and management. Much as a table of contents organizes topics in a book, the organizational structure of classification should make it easy for clinicians to find what they need. In order to organize a classification system that meets these requirements it is important to understand the cognitive conceptual structures that clinicians use in differentiating among diagnostic categories of mental and behavioural disorders. This would help us to improve the correspondence between the organization of the classification (i.e. groupings, hierarchy) and the conceptual distinctions made by clinicians. In spite of assertions by the developers of the ICD and the DSM classifications of mental disorders that usefulness to clinicians is among their highest purposes, there has been almost no direct examination of clinicians' conceptualizations as a part of their development.

Reed (2010) described the WHO operational definition of 'clinical utility' for the purposes of developing the ICD-11 classification of mental and behavioural disorders: 'the clinical utility of a

classification construct or category for mental and behavioural disorders depends on: a) its value in *communicating* (e.g., among practitioners, patients, families, administrators); b) its *implementation characteristics* in clinical practice, including its goodness of fit (i.e., accuracy of description), its ease of use and the time required to use it (i.e., feasibility); and c) its usefulness in *selecting interventions* and in making *clinical management* decisions' (p. 461). This articulation builds upon definitions offered by First et al. (2004), First (2010), and Mullins-Sweatt & Widiger (2009; Widiger & Mullins-Sweatt, 2010). Widiger, for example, has emphasized that clinical utility involves ease of using a system, communication with various parties, and planning treatment. First (2010) presented that validity and reliability are 'prerequisites of clinical utility' (p. 466), but also that improvements in validity and reliability would improve clinical utility if better diagnoses and treatment result. Improvements in other aspects of clinical utility will similarly benefit the reliability and validity of diagnostic determinations and data generated in health encounters.

In the current revision process, the WHO, in particular, views improving clinical utility as a critical orienting principle as it revises all chapters of the full ICD in order to enhance the ability of global practitioners to use the system, diagnose patients correctly, and provide the most appropriate treatments (International Advisory Group, 2011). Several commentators and groups have suggested meta-structures for organizing a classification of mental and behavioural disorders (e.g. Andrews et al., 2009; Watson, 2005; Widiger & Simonsen, 2005). However, none appear to be particularly attentive to clinical utility as viewed for the purposes described above, such that the classification organization would be as intuitive as possible for clinicians to implement in a variety of settings (First, 2009; Jablensky, 2009; Wittchen et al., 2009).

The WHO has embarked upon a series of global formative field studies to determine empirically how international clinicians conceptualize and classify mental disorders in their daily work based on their clinical experience and the ways in which they approach the clinical management of these conditions. These cognitive processes have important implications for how a classification system will be implemented in clinical practice (Flanagan et al., 2008). The studies are intended to help the WHO determine how to organize the categories in a way that is usable in different types of healthcare settings around the world. That is, how do clinicians interpret and make use of the classification system? Can disorders be organized into an intuitively useful classification construction based on empirical study of clinicians' perceptions? Which clusters of disorders

are most consistent with clinicians' conceptual maps? What is the fit of the models proposed for the ICD among international clinicians?

As part of the WHO programme of formative field studies, we conducted a quantitative evaluation of how clinicians make judgements about the inter-relationships among mental and behavioural disorders, using the worldwide web to collect data from a large-scale, international sample of psychiatrists and psychologists. We used a paired comparison procedure – a traditional psychometric technique used widely in cognitive science, marketing, and psychological research, including investigations of psychopathology – to generate a conceptual map based on clinician judgements regarding the similarity of specified diagnoses. The paired comparison methodology overcomes the tendency to apply over-learned diagnostic taxonomies to generate such judgements (i.e. the respondent could not consistently apply his or her knowledge of existing diagnostic schemes, treatment approaches, or any underlying conceptual framework to guide responses). Given the stimulus presentation strategy used in the current study, which involved the presentation of 100 randomly selected comparisons to each participant, it would be extremely difficult for participants to maintain a conscious strategy to replicate an existing classification system throughout the task. The result is an organization of the presented disorders according to clinicians' perceived relatedness of disorders, relatively free of pre-existing taxonomic biases.

People are often unaware of the dimensions they are using to make such similarity judgements; another aim of this study was to extract or 'uncover' the underlying dimensions that clinicians were using to evaluate similarity using multidimensional scaling (MDS) analysis. We then conducted analyses comparing the resulting structure to the ICD-10, DSM-IV, and proposed ICD-11 arrangements to determine the relative goodness of fit of existing and proposed diagnostic systems with how clinicians modelled mental disorders.

Method

Participants

The sample included psychiatrists and psychologists (N = 1,371) from 64 countries across all regions of the world (Table 1). Participants were solicited through the WHO international field study centres in Brazil, China, India, Mexico, Nigeria, Lebanon, Japan, Spain, and USA. Solicitation emails with a link to the web-based survey site were sent to distribution lists of regional and country psychiatric and psychological associations, lists of licensed practising psychiatrists and psychologists, reviewer lists

and editorial boards for practice-orientated journals, lists of board-certified practitioners, and practitioner presenters at international congresses. The emails contained the active link to the survey website for participants (conducted via the Qualtrics™ Research Suite).

Procedure

Participants accessed the web-based survey in either English or Spanish. The disorder labels and survey items were translated into Spanish through a collaboration of representatives from Mexico and Spain with consultation from professionals in several other Spanish-speaking countries. The initial solicitation message, introduction and informed consent, survey items and demographics questions were translated by professionals at the National Institute of Psychiatry Ramón de la Fuente Muñiz in Mexico.

Instructions stated: 'This study is intended to help us understand how clinicians classify mental disorders in their own minds, based on their clinical experience and the ways in which they approach the clinical management of these conditions.' The survey asked for the clinician's view of the similarity relationships among mental and behavioural disorders. The instructions told the participant to report about how he or she perceived the relationship of presented disorders irrespective of training in the ICD or DSM, and that there was no correct or incorrect way this task could be completed.

Stimuli and similarity ratings

Sixty disorder labels used in the study were identified through a process of international expert consensus. Because a comprehensive comparison of all possible disorders in the ICD and DSM was practically impossible, we had to limit the number of disorders presented to the participants. We conducted a pilot study to identify exemplar disorders from the ICD and DSM critical to understanding diagnostic taxonomies utilized by practitioners in clinical practice. We surveyed 25 international psychiatrists and psychologists with specific expertise in classification to select those disorders best representing the most important ones while presenting the full range of disorders in the existing classification systems. The results of this pilot survey were used to generate the 60 disorder labels used in the study (see Table 2 for the list). Disorder labels were designed so as to be understandable to both users of the ICD-10 and users of the DSM-IV. Participants were told that minor differences between disorder labels and the ones that they normally used were not intended to be meaningful.

Table 1. Descriptive demographic information on study participants.

Demographic category	M (SD)	Percentage M (SD)
Global region		
East Asia		9.4
Europe		20.1
Latin America		28.8
Middle East		4.3
Oceania		3.2
South-east Asia		15.6
Sub-Saharan Africa		5.5
USA		13.1
Gender		
Male		57.10
Female		41.76
Not reporting		1.14
Clinical profession		
Psychiatrist		63.26
Psychologist		34.56
Unreported/other		2.18
Age, education, and experience in years		
Age based on year of birth	51.93 (15.57)	
Years of speciality training	8.04 (4.64)	
Years of professional experience	15.97 (10.56)	
Primary work settings ^a		
Primary care setting		8.24
General medical setting		16.67
Mental health setting		64.87
Substance abuse setting		6.91
Private practice		44.32
University		16.67
Government agency		8.33
Non-government agency		5.78
Retired		2.08
Number of hours per week in different professional activities		
Direct care	22.61 (14.69)	
Supervision	6.27 (8.74)	
Research	7.67 (8.35)	
Teaching	5.54 (6.02)	
Administration	5.44 (8.00)	
Direct non-mental healthcare	1.83 (5.50)	
Percentages of time with different patient age groups		
Children		10.84 (18.84)
Adolescents		14.65 (15.99)
Adults		61.73 (27.38)
Elderly		12.81 (15.89)
Current use of different diagnostic systems rated 'often' or 'very often'		
ICD-10		52.37
ICD-9 or ICD-8		5.49
DSM-IV		63.45
Other system		11.65

^aNumbers add up to more than 100% because some individuals noted multiple primary work settings.

The paired comparison procedure requires a full factorial combination of all 60 disorders the number of disorder combinations ($k = 1,170$) was too large for any one clinician to complete. Consequently, we used a stimulus presentation strategy wherein each clinician was presented with a subset of 100 randomly selected pairs of disorder labels. Examples of resulting comparisons are: 'How similar are "factitious

disorder" and "nonorganic enuresis"?'; 'How similar are "intellectual disability" (mental retardation) and "autistic disorder"?'; 'How similar are "schizophrenia" and "conduct disorder"?'. The participant answered each item on a 7-point Likert scale with the end anchor labels of 'not similar at all' (1) and 'highly similar to each other' (7). The participant answered questions about his or her training and practice.

Table 2. List of disorder labels presented to the study participants.

English	Spanish	Abbreviation
Alzheimer's dementia	Demencia tipo Alzheimer	DAT
Vascular dementia	Demencia vascular	VaD
Amnestic disorder (organic)	Síndrome amnésico orgánico	AmnOrg
Delirium	Delirium (orgánico)	Delir
Mood disorder due to a general medical condition (organic)	Trastorno afectivo debido a enfermedad médica (orgánico)	Mood
Alcohol dependence	Dependencia al alcohol	Alc
Opioid dependence	Dependencia a opioides	Opiod
Cocaine dependence	Dependencia a la cocaína	Coke
Cannabinoid abuse	Abuso de cannabinoides	MJ
Abuse of volatile solvents (inhalants)	Abuso de disolventes volátiles (inhalants)	Solvent
Tobacco (nicotine) dependence	Dependencia al tabaco (nicotina)	Tobacco
Substance-induced psychotic disorder	Trastorno psicótico inducido por uso de sustancias	SubPsych
Schizophrenia	Esquizofrenia	Sz
Schizotypal disorder	Trastorno esquizotípico	Sztyp
Delusional disorder	Trastorno de ideas delirantes	Delus
Acute and transient (brief) psychotic disorder	Trastorno psicótico agudo y transitorio (breve)	ATPsyDis
Schizoaffective disorder	Trastorno esquizoafectivo	SzAff
Bipolar I disorder	Trastorno bipolar I	BiPol1
Bipolar II disorder	Trastorno bipolar II	BiPol2
Depressive disorder (major)	Trastorno depresivo (mayor)	DepDis
Cyclothymia	Ciclotimia	Cyclo
Dysthymia	Distimia	Dysth
Panic disorder	Trastorno de pánico	Panic
Social phobia	Fobia social	SocPhob
Generalized anxiety disorder	Trastorno de ansiedad generalizada	GAD
Mixed anxiety and depressive disorder	Trastorno mixto ansioso-depresivo	AnxDep
Obsessive-compulsive disorder	Trastorno obsesivo-compulsivo	OCD
Post-traumatic stress disorder	Trastorno por estrés postraumático	PTSD
Adjustment disorders	Trastornos de adaptación	Adj
Dissociative disorders	Trastornos disociativos	Dissoc
Conversion disorders	Trastornos conversivos	Convert
Somatization disorder	Trastorno por somatización	Somat
Hypochondriacal disorder	Hipocondría	Hypocho
Persistent somatoform pain disorder	Trastorno de dolor persistente somatomorfo	Pain
Body dysmorphic disorder	Trastorno dismórfico corporal	Dysmorph
Anorexia nervosa	Anorexia nervosa	AnorX
Primary (nonorganic) insomnia	Insomnio primario (no orgánico)	Insom
Sexual dysfunction	Disfunción sexual (no orgánica)	SxDys
Abuse of non-dependence producing substances (e.g. steroids, hormones)	Abuso de sustancias que no producen dependencia (ej: esteroides, hormonas)	NonDepAb
Paranoid personality disorder	Trastorno paranoide de la personalidad	ParPD
Antisocial (dissocial) personality disorder	Trastorno antisocial de la personalidad	ASP
Borderline personality disorder	Trastorno límite ('borderline') de la personalidad	BPD
Dependent personality disorder	Trastorno dependiente de la personalidad	DepPD
Pathological gambling	Juego patológico	Gamble
Intermittent explosive disorder	Trastorno explosivo intermitente	Explos
Paraphilias	Parafilias	Parphil
Gender identity disorder	Trastorno de identidad de género	GendID
Factitious disorder	Trastorno facticio	Factit
Intellectual disability (mental retardation)	Retraso mental	MR
Specific developmental disorders of speech and language	Trastornos específicos del desarrollo del habla y del lenguaje	DDLang
Specific developmental disorders of scholastic skills	Trastornos específicos del aprendizaje escolar	DDSchol
Autistic disorder	Trastorno autista	Aut
Asperger's syndrome	Síndrome de Asperger	Asperg
Attention deficit-hyperactivity (hyperkinetic) disorder	Trastorno (hipercinético) por déficit de atención con hiperactividad	ADHD
Conduct disorder	Trastorno de conducta disocial	Conduct
Oppositional defiant disorder	Trastorno oposicionista desafiante	OpDefD
Childhood separation anxiety disorder	Trastorno de ansiedad de separación en la infancia	ChildSep
Reactive attachment disorder	Trastorno reactivo de la vinculación en la infancia	ReactAtt
Tic disorders	Trastornos por tics	Tic
Non-organic enuresis	Enuresis (no orgánica)	Enuresis

Statistical methods

In the paired comparison methodology, proximity (i.e. similarity) data are used to obtain the average conceptual distance over many comparisons between each of the presented stimuli. Proximity data are ideal for creating perceptual maps using MDS to represent stimuli in coordinate space based on the conceptual similarity of the stimuli (Kruskal & Wish, 1977). This is analogous to, though more complex than, generating a map of a country based on knowledge of the distances between every town, village, or city in the country and every other one. Through an iterative process, MDS recovers a set of parsimonious dimensions used by participants to make complex judgements, and then plots those data in a perceptual map that represents the subjective interrelationships among stimuli based on those dimensions. MDS has been used extensively in a variety of studies related to clinical and diagnostic issues and to classification (e.g. Egli et al., 2006; Lease et al., 2003; Treat et al., 2002).

Individual differences multidimensional scaling (INDSCAL) is an extension of general MDS techniques that preserves individual differences. Based on the possibility that individual participants (or groups of people) may perceive given stimuli differently, INDSCAL also assumes that differences between individuals correspond to differences in the dimensional salience along which stimuli may be classified. That is, individuals are thought to use the same set of dimensions to make their ratings, but to different extents. One major advantage of this approach is that it enables between-subject comparisons, which made the INDSCAL model ideal for exploring differences between groups of participants (e.g. by country income, language and profession) in the present study. Because participants did not rate all possible disorder comparisons, matrices used in INDSCAL were populated by randomly selecting data from participants until the 60×60 disorder matrix was full while preserving some level of individual difference variability. This process was completed for each subgroup that might be compared. For instance, when participants who participated in Spanish were compared with those who participated in English, each paired comparison matrix contained a hierarchical random sample of either English speaking or Spanish speaking clinicians. As a result, the matrices used in these analyses represent clinician-specific individual differences, without requiring that each clinician complete all 1,770 comparisons.

Traditional inter-rater statistics (Cohen's kappa, Pearson's r , and Spearman's rho) were used to test the level of agreement among clinician ratings (homogeneity) and with the structure of existing classification

systems (ICD-10 and DSM-IV) and the proposed structure for ICD-11.

Results

Overall structure of the classification

Similarity ratings from the 1,371 clinicians (psychiatrists and psychologists) were submitted to the INDSCAL procedure in SAS version 9.2 (PROC MDS /COEFF = DIAGONAL). Ordinal level transformations forced the software to resolve all ties produced in the transformation. The data were configured in a two-mode, three-way proximities matrix to capture stimuli similarity across all participants and by each participant. INDSCAL produces a set of dimensions which describe the multidimensional space representing the distance of each disorder from every other disorder as determined by clinician ratings. To evaluate the overall dimensional structure of the data, three methods were used to appraise which of six possible MDS solutions (including one through six dimensions, respectively) had the best fit: (1) interpretability of the configuration, (2) scree plots of the relative change in goodness of fit (proportion of variance explained R^2), and (3) the relative change in stress (an MDS measure for badness of fit). All methods converged, indicating that a three-dimensional solution was the most parsimonious and interpretable configuration of these clinician judgement data. The pattern of relatively smaller change values above the second and third dimensions indicates that the three-dimensional solution fits these data best. The correlation among clinician judgements for this dimensional attribute was also high ($r = 0.72$). Although two- and four-dimensional solutions were also considered, the configurations of these alternative dimensional solutions were not interpretable. The use of the term 'dimension' in this context refers to the range on a continuum for the characteristics articulated by the dimension names. Dimension as used here does not relate to the literature on categorical versus dimensional models of classification (Widiger & Mullins-Sweatt, 2010).

Empirical interpretation of the three-dimensional solution dimensions was achieved by identifying exemplars of each pole and conceptually describing the transition across disorders along the dimension. Dimension 1 was interpreted as reflecting an 'internalizing versus externalizing' trend. For example, diagnostic categories that located towards the internalizing end of this dimension included somatization disorder, conversion disorder, and hypochondriacal disorder. Diagnostic categories located towards the externalizing end included antisocial personality disorder, abuse of volatile solvents, and intermittent

explosive disorder. Dimension 2 was interpreted as ‘developmental versus adult onset’. Diagnostic categories located towards the developmental end of this dimension included specific developmental disorders of speech and language, autistic disorder, and Asperger’s syndrome. Diagnostic categories located towards the adult onset end of this dimension included cyclothymia, bipolar disorders I and II, and some substance-related diagnoses such as opioid and tobacco dependence. Dimension 3 was interpreted as ‘functional versus organic’ syndromes. Diagnostic labels judged as more towards the functional end of the dimension included dependent personality disorder, sexual dysfunction, and reactive attachment disorder. Diagnostic labels judged as more organic included vascular and Alzheimer’s dementias, amnestic disorder, and delirium. It should be noted that individual disorders are more or less described by each dimension; some disorders could be relatively neutral on one dimension but highly characterized by another. That is, not all of the dimensions are necessarily as relevant to some disorders as others.

Plots of the location of the disorders on each combination of two dimensions are shown in Figs. 1, 2 and 3. (Disorder labels are provided in Table 2.) We caution the reader that Figs. 1 to 3 are two-dimensional representations of a three-dimensional solution; each view is therefore missing important information.

To gain insight into the complex relationship among these three dimensions of diagnostic judgements, an example is presented to demonstrate the mapping of diagnostic concepts using the three dimensions simultaneously. In Fig. 4, disorders related to substance use/addiction and those with primary psychotic symptoms are depicted in three-dimensional space. All disorders are towards the ‘externalizing’ side of dimension 1 (values higher than zero) with the exception of delusional disorder, which is just slightly towards the ‘internalizing’ side. Schizophrenia and schizotypal disorder are viewed as relatively more ‘developmental’ in nature than most of the other disorders. Substance-induced psychotic disorder is considered to be trending towards ‘organic’ as are most of the psychotic disorders. However, when all three dimensions are taken into account simultaneously, there is a clear clustering of these two groups of disorders, and a clear demarcation between them formed through the interaction of the dimensions. As might be expected, substance-induced psychotic disorder is placed at the intersection of the two diagnostic groups. Interestingly, paranoid personality disorder, along with schizotypal disorder, is placed nearer primary psychotic disorders, rather than near other personality disorders (e.g. see Fig. 2). Thus, a series of independent and random comparisons across a large international sample of clinicians yielded a coherent set of groups that correspond meaningfully to each other.

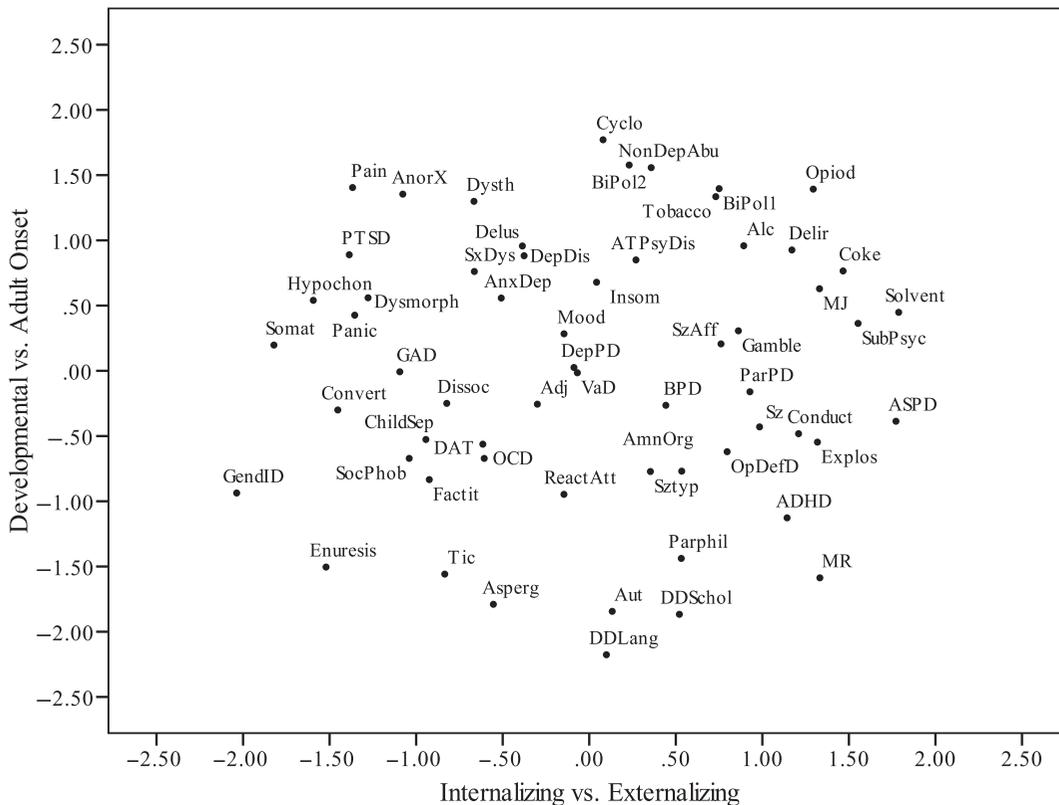


Fig. 1. Clinicians’ organization of disorders by dimensions 1 and 2. See Table 2 for abbreviations of disorders depicted.

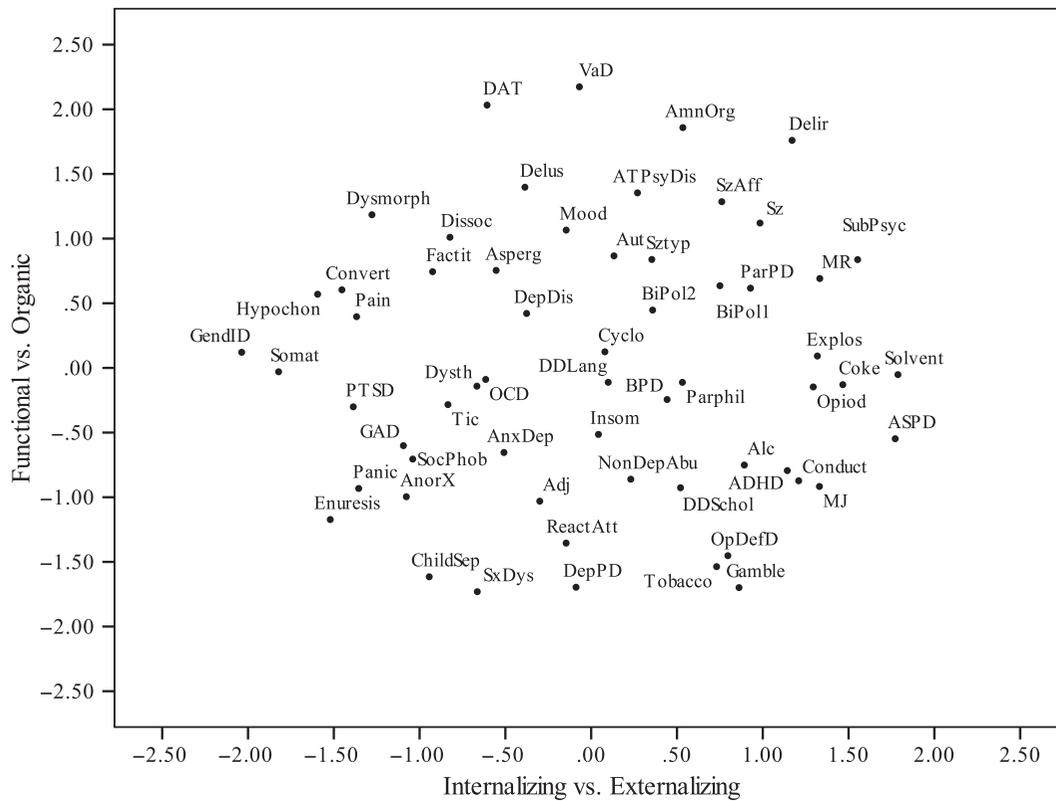


Fig. 2. Clinicians' organization of disorders by dimensions 1 and 3. See Table 2 for abbreviations of disorders depicted.

Consistency across clinician groups

Consistency across clinician groups was examined using comparisons among standard Pearson correlation coefficients of averaged similarity ratings for subgroups of clinicians within the total sample. The conceptual map described above was highly consistent across clinicians. First, English speaking participants' ratings were remarkably similar to those of Spanish speaking participants ($r = 0.83$). Second, psychologists' ratings were very close to those of psychiatrists ($r = 0.81$). Third, when participants were grouped by country income level according to the World Bank country income classification (World Bank, 2010), clinicians' solutions remained highly consistent (low and lower-middle to middle $r = 0.81$, low and lower-middle to high $r = 0.78$, middle to high $r = 0.77$). For the purpose of this analysis, countries with low and lower-middle incomes were combined due to the limited number of participants from low-income countries within the sample.

Comparison across classification systems (ICD-10 versus DSM-IV versus ICD-11)

We further examined the structure of clinicians' implicit classifications of mental disorders by comparing them to existing classification systems (ICD-10 and DSM-IV), and to the proposed structure of the

ICD-11 classification of mental and behavioural disorders. The structure for ICD-11 proposed at the time of these analyses is shown in Table 3. ICD-11 disorder groupings have not been finalized and are subject to modification, including based on the results of this study. For each classification system we determined whether a disorder pair was placed in the same grouping within each of the three classification systems, in which case the pair received a score of 1, or whether they were placed in a separate category, in which case the pair received a score of 0. The consistency between the classification systems (as measured by kappa) is a useful baseline for comparing the consistency of the clinicians' structure to each system. In order to standardize the measurement scheme to comparable units, we first treated the clinicians' ratings as ordinal rather than continuous, thereby calculating Spearman's rho as a measure of consistency with the classification systems. Next, we treated clinicians' ratings as a dichotomous variable, thereby calculating kappa values for a measure of consistency. The results of the two analyses were identical; therefore we present only one set of values (see Table 4). The structure generated by clinicians reaches a level of agreement with each system that is roughly equivalent to the agreement among the systems. However, generally speaking, the agreement among classification systems and of clinicians'

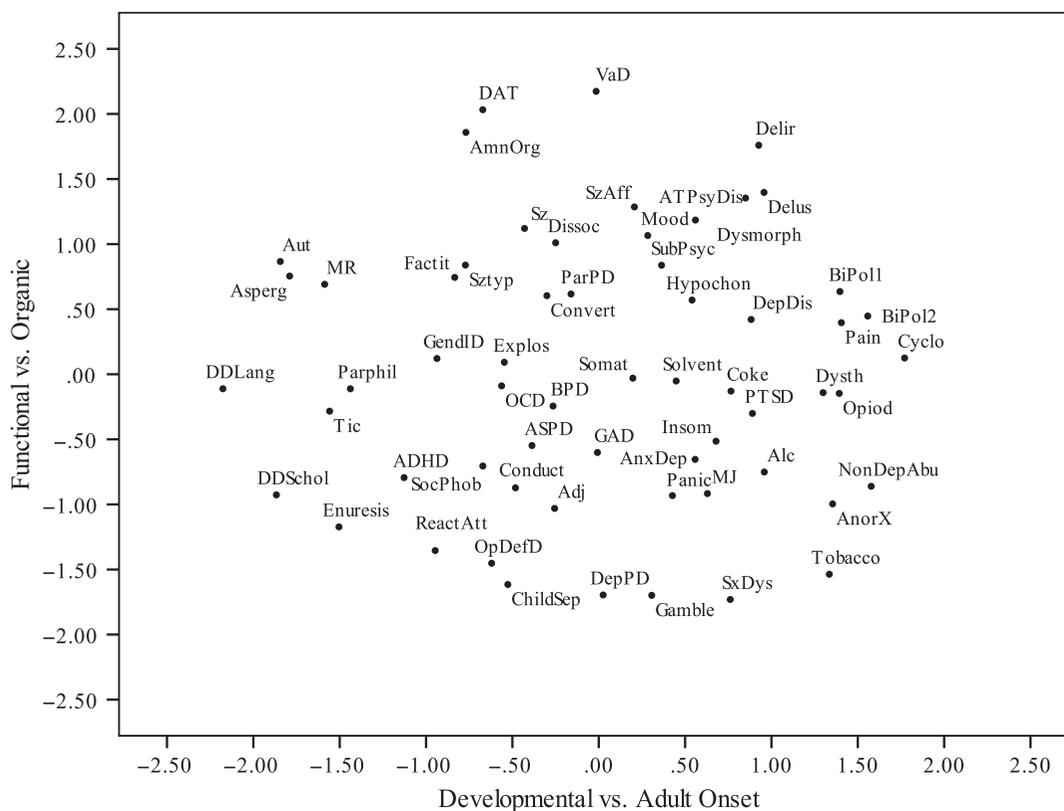


Fig. 3. Clinicians' organization of disorders by dimensions 2 and 3. See Table 2 for abbreviations of disorders depicted.

consensual structure to each classification system was moderate.

Nonetheless, the proposed ICD-11 appears to represent an improvement of fit with clinicians' inherent organization as evidenced by higher kappa values. To test the incremental association between ICD versions and clinicians' similarity ratings more systematically, we conducted a hierarchical regression using the continuous version of clinicians' ratings as the dependent measure predicted by the structure of the successive ICD versions. First, the prediction of clinicians' ratings by ICD-10 and ICD-11 was significant; $F(2,1767) = 464.92$, $p < 0.0001$, $R^2 = 0.34$. The ICD-10 predicted an initial 23% of the variance (semi-partial $R^2 = 0.23$) with the revision of the ICD-11 predicting an additional 11% (semi-partial $R^2 = 0.11$). Thus, the proposed organization of the ICD-11 represents an improvement of fit with clinicians' implicit model of psychopathology, though considerable divergence still exists between the clinicians' consensual model and the proposed ICD-11 structure.

Discussion

The present study found that clinicians view mental and behavioural disorders along three distinctive dimensions based on which they consistently categorize

related disorders. Meaningful clusters of disorders and the relationships among these clusters can be drawn from the similarity ratings of the disorders. We have labelled for purposes of communication the three dimensions that clinicians used as a basis for their ratings: (1) internalizing versus externalizing; (2) developmental versus adult onset; and (3) functional versus organic. Although other labels might also be applied to these dimensions, these appear to be fairly descriptive of their characteristics. These dimensions are analogous to the north–south and east–west axes on a map; we emphasize that they do not represent diagnostic categories or groupings. That is, we are not suggesting that the classification system should be organized by grouping the disorders according to the labels for the extreme ends of these dimensions (e.g. 'externalizing disorders', 'organic disorders') and creating categorical entities or groupings on that basis. Instead, having this map and the dimensions enhances the ability to consider related disorders and clusters of disorders in the data, along the lines of what is shown in Fig. 4, and also to evaluate the extent to which any proposed organization is more or less consistent with the clinicians' consensual model. This might include placement of a disorder closer to another in the organizational structure to help clinicians find and correctly apply the diagnosis.

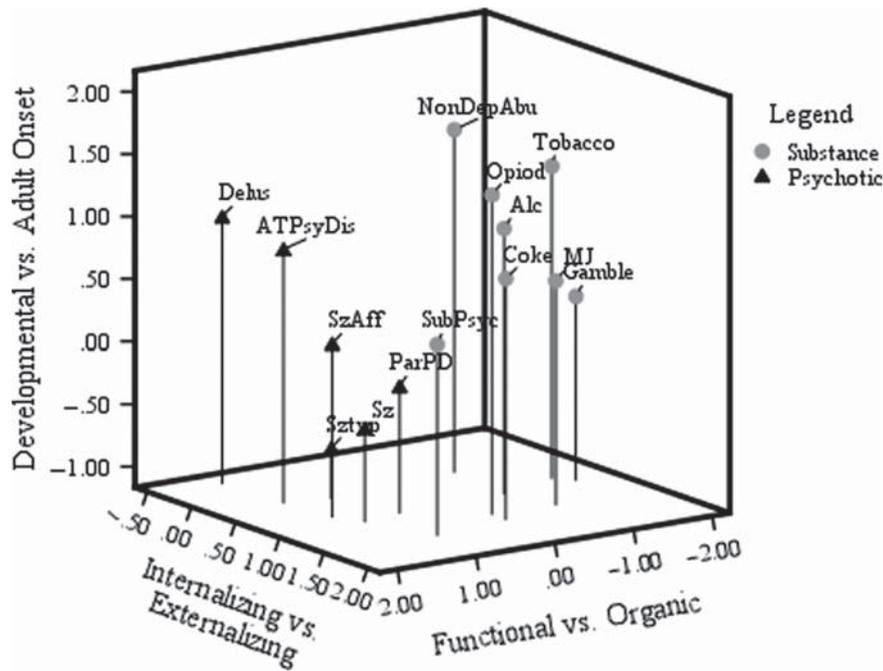


Fig. 4. Three-dimensional organization of the substance abuse and psychotic disorders. See Table 2 for abbreviations of disorders depicted.

These models obtained from clinicians are not a replication of either the ICD-10 or the DSM-IV; participating clinicians did not simply reproduce the relationships that would have been dictated by their training or daily use of either system. Which diagnostic system clinicians actually used in their clinical practice did not have an effect on the clinicians' conceptual map. The methodology removed influences of prior training and experience with a classification system to tap a more universal set of perceptions of

relatedness among the disorders. Thus, the dimensions represent more directly the underlying connections and constructs that clinicians inherently hold in their consensual taxonomy. Results indicate that clinicians simultaneously use the three dimensions in conceptualizing the degree of similarity between disorders. That is, they weigh multiple types of information in making this decision. The fact that existing classifications allow for only a single grouping principle rather than the simultaneous consideration of multiple dimensions is clearly one reason for the only modest ability of the ICD and DSM groupings to predict the placement of disorders in the clinicians' consensual conceptual model.

Table 3. Proposed ICD-11 disorder groupings used in analyses of consistency between classification systems and clinicians' ratings.

Major proposed disorder groupings for disorders used in the present study

- Neurodevelopmental disorders
- Schizophrenia spectrum and primary psychotic disorders
- Bipolar and related disorders
- Depressive disorders
- Anxiety and fear-related disorders
- Disorders specifically associated with stress
- Dissociative disorders
- Bodily distress disorders
- Obsessive-compulsive and related disorders
- Feeding and eating disorders
- Elimination disorders
- Sleep disorders
- Sexual disorders
- Substance-related and addictive disorders
- Neurocognitive disorders
- Personality disorders
- Mental and behavioural disorders associated with disorders or diseases classified elsewhere
- Other mental and behavioural disorders

Clinicians' conceptualizations of the relationships among mental disorders are rational and highly stable across clinician characteristics, regardless of profession, language, and country income level. Thus, a commonality of conceptualization based on clinical experience and clinician expertise appears to be driving the similarity across global clinicians. These results suggest that differences between clinicians' judgements and classification systems are not due to idiosyncratic clinician error, but rather to systematic

Table 4. Kappa values for the consistency between classification systems and clinicians' ratings.

	ICD-11	DSM-IV	Clinicians
ICD-10	0.41	0.47	0.42
ICD-11	—	0.54	0.51
DSM-IV	—	—	0.42

and stable perceptions. Two examples are worth highlighting: clinicians locate ADHD in the conceptual model between disruptive behaviour disorders (e.g. conduct disorder, oppositional defiant disorder) and neurodevelopmental disorders (e.g. specific developmental disorders, mental retardation). They also locate substance-induced psychosis as between substance use disorders and psychotic disorders. These placements are logical and automatic (given the methodology) to international clinicians based on their clinical experience, regardless of what research might indicate about the commonality of their aetiology, neurobiological correlates, symptom presentation, or appropriate treatments. These placements also demonstrate how forcing an inherently dimensional conceptual model into a categorical classification system results in a loss of information. These results are consistent with Flanagan and Blashfield's (2007) observation that 'clinicians, with their years of experience dealing with the ambiguities, the emotionality, the chaos of psychopathology, have formed complex conceptual systems to try to understand the mysteries of abnormal behavior' (pp. 285–286).

In some ways, taking these results seriously turns much of the existing discourse about classification on its head. Rather than simply assuming that clinician ignorance accounts for their failure to conform to the ostensibly scientific gold standard of classification, we may also observe that existing (and proposed) classifications continue to be relatively poor reflections of a rich and consensual conceptual model held by global clinicians based on their clinical experience and knowledge, about which there appears to be considerably more agreement among clinicians than there is among current classification systems. Viewed in this way, understanding the clinicians' conceptual map of mental disorders has major implications for the design and implementation of a classification system that clinicians will be expected to use. For example, consider body dysphoric disorder and obsessive-compulsive disorder. In the present study, the clinicians did not perceive them as similar, so consequently they are not close together on clinician's conceptual maps (see Figs. 1–3), although there is some research evidence of relatedness, and DSM-5 has proposed grouping them together (see www.dsm5.org).

We are not arguing that clinician judgements should be followed when science tells us their conceptualizations are incorrect. In developing the ICD-11 it may be determined that the weight of the scientific evidence supports grouping body dysphoric disorder and obsessive-compulsive disorder together in the classification system, in spite of the fact that clinicians do not perceive them as similar. This study suggests that during implementation of the ICD-11 it would

be important to provide special emphasis on this particular grouping in educational efforts aimed at professionals at all levels of preparation. For example, education might include a description of the scientific justification for the grouping and a clinical rationale that would help clinicians understand what the clinical and management implications are of conceptualizing these disorders as a part of the same group. On the other hand, if it were determined that grouping these disorders together was scientifically justifiable (e.g. based on genetic evidence), but had no implications in terms of treatment or management, an alternate decision might be made to cluster them in a different way. One could view the structure of the classification as representing a 'user interface' or table of contents, rather than the most accurate possible reflection of the current state of our scientific knowledge. For the clinical version of the classification, the disorders could be grouped in a way that is as parallel as possible to clinicians' conceptualizations in order to assist users in finding the category and applying the appropriate diagnosis. A research version of the classification might be organized differently.

The dimensional organization elicited from clinicians does not preclude the use of scientifically based classification principles, but describes an approach for how to organize the disorders so that clinicians can easily find what they need in order to make appropriate diagnoses. In fact, the dimensions found in this study are not in any way inconsistent with current scientific evidence. For example, internalizing-externalizing was the first dimension to emerge in the MDS analysis, and current literature suggests this to be the most important empirically derived dimension in studies of the structure of common mental disorders (e.g. Krueger, 1999; Krueger et al., 2007). It would also be difficult to mount a scientific argument against developmental period of onset and the extent to which a disorder involves brain dysfunction, illness, or injury as organizing principles in a classification of mental disorders.

Existing classification systems (ICD-10 and DSM-IV) do not fully reflect how clinicians think about the mental disorders they encounter in practice, as evidenced by only moderate correlations of clinician perceptions to current diagnostic systems. The proposed meta-structure for ICD-11 corresponds better to the clinicians' conceptual map of mental disorders than either ICD-10 or DSM-IV; the increase in explained variance represents an incremental improvement over existing systems. A modest improvement is perhaps a reasonable expectation for the current ICD revision process; given the current state of knowledge, classification systems will remain, at best, an imperfect representation of clinicians' conceptualizations of the mental and behavioural disorders they work with every day.

Because current diagnostic systems are often seen as lacking in clinical utility, they have often been frustrating to clinicians, and diagnostic guidelines have been poorly applied as a result. A multidimensional structure for organizing disorders, such as the one observed in this study, could be used to help all healthcare providers find the categories they need to use as quickly and efficiently as possible. This consideration is especially important for the many primary care providers who are not trained extensively in diagnosis of mental and behavioural disorders.

Any system must meet multiple needs. In order to meet public health needs, an important goal of the current revision should be to improve the classifications' suitability for implementation in clinical settings (International Advisory Group, 2011; Reed, 2010). Both the ICD-10 and the DSM-IV assert that clinical utility was of major importance in their development, but in fact, the development of current classification systems involved no systematic investigation of how mental disorders categories could be organized to better suit clinical needs (First, 2010; Reed, 2010). At the time of the approval of ICD-10 in 1990, and of the DSM-IV in 1994, the assumption was that the next version of classification systems for mental and behavioural disorders would be strongly influenced by research into biological bases of mental disorders, including empirical advancements in genetics, neurosciences, assessment instruments, and treatments (Andrews et al., 2009; Hyman, 2007; Insel et al., 2010). Unfortunately for the current revisions, while there have indeed been dramatic advances in these areas, they have not produced specific findings based on which to change classification systems substantially (Hyman, 2007; Reed, 2010). If current revisions to classification systems are not going to dramatically alter the structure and descriptions of mental disorders based on biopsychosocial data to improve their validity, an appropriate focus is to improve their clinical utility in order to facilitate identification and treatment of mental disorders to reduce the disease burden associated with them.

Strengths and limitations

Particular strengths of this project include the very large international sample of psychiatrists and psychologists producing highly consistent results. The paired comparison approach and statistical analyses provide a robust methodology to examine the relationships among disorders as clinicians conceptualize them. We note as a possible limitation that the present study measured clinicians' perceptions of disorder constructs; it was not based on patient diagnoses or statistical relationships among presenting symptoms. The latter was not our purpose; we set out to develop information useful in mapping how

a classification system might be organized to enhance the ability of global practitioners to use the system globally, diagnose patients correctly, and provide the most appropriate treatments (International Advisory Group, 2011). This issue is especially important for primary care providers who are often the first responders to mental health problems but who do not receive extensive training in the diagnostic systems (Reed, 2010; Reed et al., 2009).

Implications

As the WHO ICD Working Groups deliberate how to group mental and behavioural disorders in the new ICD-11 classification, they must consider the needs of the various constituencies including specialist clinicians, non-specialist primary care practitioners, and researchers. This understanding of how clinicians form groupings of disorders can help create a classification system that has greater utility for real-life clinical work. Busy clinicians, faced with the need to make diagnostic and patient care decisions under tremendous time pressures, will turn to the place in the classification system where they would most naturally think to find the disorder and its description (without engaging in intense debate over disorder aetiology and clinical treatment or researchers' fine discriminations), and may not continue to look if they do not find it there. Clinician judgements can be used to improve clinical utility of ICD-11, based on a scientific approach, without sacrificing validity. Examining clinicians' views of mental and behavioural disorders and the relationships among them is an important and necessary process to developing a useful classification system, because ultimately the system must make sense to the clinician who implements it (Flanagan et al., 2008). This approach fulfils the clinical utility orientation that WHO has established for the ICD revision process.

Declaration of interest: Pratap Sharan received a grant from Eli Lilly (India) unrelated to this project; Jair de Jesus Mari received speaking fees from Astra Zeneca, Eli Lilly, and Janssen unrelated to this project. 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). The University of Kansas (USA) provided logistical assistance in the conduct of this study. 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. Geoffrey Reed is Senior Project Officer, Revision of ICD-10 Mental and Behavioural Disorders, 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 the WHO. The WHO Research Ethics Review Committee and the Human Subjects Committee of the University of Kansas, USA, reviewed and approved all procedures. The authors alone are responsible for the content and writing of the paper.

References

- Andrews, G., Anderson, T., Slade, T. & Sunderland, M. (2008). The classification of anxiety and depression: Problems and solutions. *Depression and Anxiety, 25*, 274–281.
- Andrews, G., Goldberg, D.P., Krueger, R.F., Carpenter, W.T., Jr., Hyman, S.E., Sachdev, P. & Pine, D. (2009). Exploring the feasibility of a meta-structure for DSM-V and ICD-11: Could it improve utility and validity? *Psychological Medicine, 39*, 1993–2000.
- Egli, S., Schlatter, K., Streule, R. & Lage, D. (2006). A structure-based expert model of the ICD-10 mental disorders. *Psychopathology, 39*, 1–9.
- First, M.B. (2009). Reorganizing the diagnostic grouping in DSM-IV 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. (2007). Should clinicians' views of mental illness influence the DSM? *Philosophy, Psychiatry and Psychology, 14*, 285–287.
- Flanagan, E.H. & Blashfield, R.K. (2010). Increasing clinical utility by aligning the DSM and ICD with clinicians' conceptualizations. *Professional Psychology: Research and Practice, 41*, 474–481.
- 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.
- Hyman, S.E. (2007). Can neuroscience be integrated into the DSM-V? *Nature Reviews Neuroscience, 8*, 725–732.
- Hyman, S.E. (2010). The diagnosis of mental disorders: The problem of reification. *Annual Review of Clinical Psychology, 6*, 155–179.
- Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D.S., Quinn, K., ... Wang, P. (2010). Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. *American Journal of Psychiatry, 167*, 748–751.
- International Advisory Group (for the Revision of ICD-10 Mental and Behavioral 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. (1999). The structure of common mental disorders. *Archives of General Psychiatry, 56*, 921–926.
- Krueger, R.F., Markon, K.E., Patrick, C.J., Benning, S.D. & Kramer, M. (2007). Linking antisocial behavior, substance use, and personality: An integrative quantitative model of the adult externalizing spectrum. *Journal of Abnormal Psychology, 116*, 645–666.
- Kruskal, J.B. & Wish, M. (1977). *Multidimensional Scaling*. Beverly Hills, CA: Sage.
- Lease, A.M., McFall, R.M., Treat, T.A. & Viken, R.J. (2003). Assessing children's representation of their peer group using a multidimensional scaling technique. *Journal of Social and Personal Relationships, 20*, 707–728.
- 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.
- Reed, G.M., Sharan, P. & Saxena, S. (2009). Reducing disease burden through the revision of the ICD-10 mental and behavioural codes. *National Medical Journal of India, 22*, 285–288.
- Treat, T.A., McFall, R.M., Viken, R.J., Nosofsky, R.M., MacKay, D.B. & Kuschke, J.K. (2002). Assessing clinically relevant perceptual organization with multidimensional scaling techniques. *Psychological Assessment, 14*, 239–252.
- Watson, D. (2005). Subtypes, specifiers, epicycles, and eccentrics: Toward a more parsimonious taxonomy of psychopathology. *Clinical Psychology: Science and Practice, 10*, 233–238.
- 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.
- Widiger, T.A. & Simonsen, E. (2005). Alternative dimensional models of personality disorder: Finding a common ground. *Journal of Personality Disorders, 19*, 110–130.
- Widiger, T.A. & Mullins-Sweatt, S.N. (2010). Clinical utility of a dimensional model of personality disorder. *Professional Psychology: Research and Practice, 41*, 488–494.
- World Bank (2010). Country income groups. Available at: <http://data.worldbank.org/about/country-classifications>

Copyright of International Review of Psychiatry is the property of Routledge and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.