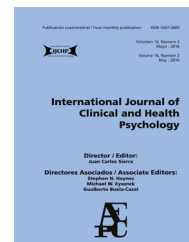




International Journal of Clinical and Health Psychology

www.elsevier.es/ijchp



ORIGINAL ARTICLE

ICD-11 guidelines for psychotic, mood, anxiety and stress-related disorders in Mexico: Clinical utility and reliability



María Elena Medina-Mora^{a,*}, Rebeca Robles^a, Tahilia J. Rebello^b,
Tecelli Domínguez^c, Nicolás Martínez^a, Francisco Juárez^a, Pratap Sharan^d,
Geoffrey M. Reed^{a,b}

^a Instituto Nacional de Psiquiatría Ramón de la Fuente, Mexico

^b Global Mental Health Program, Columbia University College of Physicians and Surgeons, USA

^c Cátedra CONACYT (Consejo Nacional de Ciencia y Tecnología), Mexico

^d All India Institute of Medical Sciences, Ansari Nagar, India

Received 28 June 2018; accepted 24 September 2018

Available online 15 October 2018

KEYWORDS

ICD-11;
Diagnostic guidelines;
Mental disorders;
Reliability;
Clinical utility;
Instrumental study

Abstract *Background/Objective:* The World Health Organization's diagnostic guidelines for ICD-11 mental and behavioural disorders must be tested in clinical settings around the world to ensure that they are clinically useful and genuinely global. The objective is evaluate the inter-rater reliability and clinical utility of ICD-11 guidelines for psychotic, mood, anxiety- and stress-related disorders in Mexican patients. *Method:* Adult volunteers exhibiting the selected symptoms were referred from the pre-consultation unit of a public psychiatric hospital to an interview by a pair of clinicians, who subsequently assigned independent diagnoses and evaluated the clinical utility of the diagnostic guidelines as applied to each particular case, on the basis of a scale developed for this purpose. *Results:* 23 clinicians evaluated 153 patients. Kappa scores were strong for psychotic disorders (.83), moderate for stress-related (.77) and mood disorders (.60) and weak for anxiety and fear-related disorders (.43). A high proportion of clinicians considered all diagnostic guidelines to be quite to extremely useful as applied to their patients. *Conclusions:* ICD-11 guidelines for psychotic, stress-related and mood disorders allow adequate inter-rater consistency among Mexican clinicians, who also considered them as clinical useful tools.

© 2018 Asociación Española de Psicología Conductual. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

* Corresponding author: Dirección General, Instituto Nacional de Psiquiatría Ramón de la Fuente Muñiz, Calzada México-Xochimilco 101, Ciudad de México, 1437, Mexico

E-mail address: medinam@imp.edu.mx (M.E. Medina-Mora).

<https://doi.org/10.1016/j.ijchp.2018.09.003>

1697-2600/© 2018 Asociación Española de Psicología Conductual. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

PALABRAS CLAVE

Guías diagnósticas;
trastornos mentales;
fiabilidad;
utilidad clínica;
estudio instrumental

Guías CIE-11 para trastornos psicóticos, afectivos, de ansiedad y estrés en México: utilidad clínica y fiabilidad

Resumen *Antecedentes/Objetivo:* Las guías diagnósticas CIE-11 para trastornos mentales y del comportamiento de la Organización Mundial de la Salud deben ser evaluadas en pacientes reales alrededor del mundo a fin de asegurar que son clínicamente útiles y genuinamente globales. Se evalúa la consistencia inter-evaluadores y la utilidad clínica de las guías para los trastornos psicóticos, afectivos, de ansiedad y relacionados con el estrés en pacientes mexicanos. *Método:* Voluntarios con síntomas psicóticos, afectivos, de ansiedad o relacionados con el estrés derivados de una unidad de pre-consulta de un hospital psiquiátrico, para una entrevista con una pareja de clínicos, quienes posteriormente asignaron diagnósticos de manera independiente y evaluaron la utilidad clínica de las guías aplicadas a cada caso en particular, con base en una escala desarrollada para este propósito. *Resultados:* 23 clínicos evaluaron 153 pacientes. Los coeficientes Kappa fueron fuertes para trastornos psicóticos (0,83), moderados para los relacionados con el estrés (0,77) y afectivos (0,60), y débiles para los de ansiedad y relacionados con el miedo (0,43). Una alta proporción de clínicos consideró que las guías eran bastante o extremadamente útiles. *Conclusiones:* Las guías CIE-11 para dichos trastornos permiten una adecuada consistencia inter-evaluadores en clínicos mexicanos, quienes les consideran herramientas clínicamente útiles.

© 2018 Asociación Española de Psicología Conductual. Publicado por Elsevier España, S.L.U. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

The World Health Organization's (WHO) diagnostic guidelines for the Mental and Behavioral Disorders chapter of the Eleventh Revision of the International Classification of Disease and Related Health Problems (ICD-11) were developed by WHO-appointed expert Working Groups (WG); the process used to develop guidelines has been described in detail elsewhere (First, Reed, Saxena, & Hyman, 2015). The guidelines were tested in previous internet case-control studies (Keeley, Reed, Roberts, Evans, Medina-Mora et al., 2016) designed to evaluate the impact of changes in the classification from ICD-10 to ICD-11 in diagnostic decisions. The guidelines were subsequently modified on the basis of the results of these studies, with the WHO expert WG suggesting the modifications and overseeing the process. The next step was to test the guidelines and their impact on decision-making in real settings in order to confirm that they do in fact lead to improvements in diagnostic practice in clinical settings around the world.

Having reliable guidelines with a high level of clinical utility (Reed, 2010) supports WHO's overarching aim of reducing the disease burden of mental and behavioral disorders (International Advisory Group for the Revision of ICD-10 Mental and Behavioural Disorders, 2011). For the guidelines to be considered clinically useful, they should be accurately and easily used by practitioners (Reed et al., 2015); and they broad application in different countries helps to show that they are genuinely global (Reed et al., 2018).

In Latin America, implementation of the ICD-11 diagnostic guidelines will take place in a particular context. In this region, years lived with disability due to depression range from 10.5% in Paraguay to 7.5% in Guatemala and Venezuela, and for anxiety disorders from 7.6% in Paraguay to 4% in Mexico (World Health Organization WHO, 2017). Recent decades have seen an increase in violence in many countries, two

(Honduras and Venezuela) of which are ranked as having the first and second highest homicide rates worldwide (United Nations Office on Drugs and Crime UNODC, 2013). Violence is linked to both mental disorders and suicide (Benjet, Borges, G., & Medina-Mora, 2010; Liu et al., 2017). In countries in the region included in the World Mental Health Survey, PTSD ranges from 4.9% in Medellín, Colombia to 0.8% in Peru (Bromet et al., 2017). The treatment gap between those who need services and those who receive them is high, amounting to 73% of those diagnosed with mental disorders (Pan American Health Organization PAHO, 2013).

In Mexico, according to the latest Psychiatric Epidemiology Survey, approximately one in four adults (between the ages of 18 and 65) living in urban areas have had a mental disorder at some time in their lives, with anxiety and depression being the most common (14.3% and 9.21%, respectively) (Medina-Mora, Borges, Benjet, Lara, & Berglund, 2007), and psychosis the most disabling (Navarro et al., 2017). Prevalence rates in Mexico rank around the median among countries that are part of the World Mental Health Surveys (Kessler et al., 2007). Unfortunately, only 11% receive minimally adequate treatment; this gap is higher than what is observed in countries with similar level of development (Wang et al., 2007). This highlights the urgent need for the timely identification of cases requiring treatment.

Although insufficient alone given their limitations, diagnostic guidelines are an essential first step to identify and provide evidence-based care for patients (Craddock, & Mynors-Wallis, 2014). Nowadays, some of such limitations could be addressed in certain ways as part of the revision and improvement of a nosology system, while other would depend on the future state of understanding of the brain, particularly its higher functions. Thus, although problems of validity given that diagnoses are based on descriptive data

rather than in relation to brain function could not be easily solve by now, a more pragmatic and less rigid ICD-11 might facilitate sensible clinical diagnoses, while avoiding the exclusion of many patients that not meet strict diagnostic criteria and creates the need for multiple "comorbidity" (Craddock, & Mynors-Wallis, 2014).

This paper shows the results of the ecological studies to test the proposed ICD-11 guidelines for non-psychotic and psychotic adult patients presenting for care at a tertiary public mental health facility in Mexico. Its principal aim was to show the value of the diagnostic guidelines in informing practitioners about the specific diagnosis of their patients, their implementation characteristics (goodness of fit, ease of use and time required to apply them) and their utility in selecting interventions and making clinical management decisions (Reed, 2010). This was done by determining inter-rater consistency in diagnoses and the clinical utility of the proposed ICD-11 diagnostic guidelines for the ICD-11 groups of disorders that account for the largest share of the disease burden of mental disorders and the major proportion of service utilization in mental health settings: (1) Schizophrenia and Other Primary Psychotic Disorders; (2) Mood Disorders; (3) Anxiety and Fear-Related Disorders; and (4) Disorders Specifically Associated with Stress.

Method

This was a cross-sectional study, drawing on a sample of participants seeking mental health services in a public, specialized, mental health care setting in Mexico City, Mexico. It follows the study design developed by our international group (Reed et al., 2018) that was specifically intended to isolate the impact of the diagnostic guidelines on diagnostic assignment by clinicians (interpretation variance) rather than other sources of variability in diagnostic agreement/disagreement (e.g. information variance, observation variance). It is not intended as a test of the stability of participants' clinical presentations across time. Alternative methods, such as using independent interviews, would not control for variability in case presentations over time and information variance and would therefore be unable to provide specific information on how to improve diagnostic guidelines, the core purpose of this study. We are less interested in inter-rater reliability as a statistic and more interested in the consistency of implementation of diagnostic guidelines in circumstances where diagnostic verdicts would be the same if the guidelines were error-free.

Participants

Patients with: (1) psychotic symptoms; or (2) mood, anxiety, or stress-related symptoms without psychotic symptoms were identified by a clinician working at the outpatient psychiatric service. Identification was based on the normal intake interview performed by a second-year psychiatry resident; the intake interview is basically intended to triage patients. The information yielded by these interviews includes sociodemographic data, current reason for consultation, basic information about the course and clinical presentation of the problem, which was used to select the protocol for the patient. In the presence of psychotic

symptoms, the patient was referred to protocol 1, and in the presence of mood, anxiety or stress-related symptoms without psychotic symptoms, the patient was referred to protocol 2. We used this screening procedure to select an enriched sample of study participants likely to display the conditions that were the focus of the study (Reed et al., 2018).

After receiving a comprehensive explanation of the nature and aims of the study, and giving their written informed consent, all participants were interviewed simultaneously by two clinicians. One clinician in the pair was designated as the primary interviewer for that particular patient and the other as the observer.

Clinician raters

Clinician raters were psychiatrists, or fifth-year psychiatry residents actively engaged in clinical work (i.e., involved in the assessment or treatment of people with mental health conditions) for an average of 10 or more hours per week.

All clinician raters participated in a half-day training session on the diagnostic guidelines and study procedures. ICD-11 diagnostic guidelines for the four disorder groups included in the study were provided to participating clinicians, who were asked to read them in detail prior to the face-to-face training session. The training curriculum and materials used for the face-to-face training, developed by WHO, comprised a presentation of the innovations proposed for the ICD-11 diagnostic guidelines for each diagnostic group included and the main conceptual features of the diagnostic guidelines for each category. As part of the training, clinician raters practiced applying the diagnostic guidelines to case vignettes, and discussed the issues that arose during this process. Clinician raters were also provided with information on the study purpose, rationale, and methods, including a tutorial on how to use the Electronic Field Study System for data entry.

Procedures

The local Institutional Ethics Review Board approved all the procedures used as a part of this study, including the consent forms for both service users and clinicians. Although clinician raters had not been informed of any diagnostic formulation made by the referring clinician before conducting their diagnostic interview, they were provided with a brief clinical summary of the participant prepared by the second-year resident conducting the triage intake interview that did not include diagnoses or psychotropic medications.

During the training, clinician raters were informed that they could also review other clinical information on the patients if necessary and available (including laboratory tests and brain images), with the proviso that both clinicians should look at the same information. Clinician raters then conducted a diagnostic interview of the participant in the way they deemed most appropriate. No specific instructions were provided for the interview except that in Protocol 1 (participants with psychotic symptoms), they should ensure they assessed Schizophrenia and Other Primary Psychotic Disorder, and in Protocol 2 (participants without psychotic symptoms but with affective, anxiety- or

stress-related symptoms), they should ensure they assessed Mood Disorders, Anxiety and Fear-Related Disorders, and Disorders Specifically Associated with Stress. They were also instructed to assess any other diagnostic area appropriate to the participant's presentation, just as they would in a regular diagnostic interview. The member of the dyad designated as the interviewer for that participant conducted the interview, but the observer was allowed to ask additional questions at the end of the interview.

Clinician raters individually and autonomously entered the results of the diagnostic interview into a secure web-based electronic data capture system (the Electronic Field Studies System, developed using the Qualtrics survey platform specifically designed by the WHO Field Studies Coordination Group) for these studies (Reed et al., 2018). Clinician Raters selected up to three diagnoses they thought were applicable for the service user they had seen, or indicated that no diagnosis was warranted, and then provided diagnostic evaluation information including a thorough review of the essential features of each selected diagnostic category. This was done to ensure clinicians to include at least one of the diagnosis under study (Schizophrenia or Other Primary Psychotic Disorder in Protocol 1, and a Mood, Anxiety and Fear-Related, or Stress-Related Disorder in protocol 2), as well as the principal comorbid diagnoses within the same group of disorders or in other one.

In addition, clinician raters provided data on the severity of the service user's symptoms and their functional status, and answered questions about the clinical utility of the ICD-11 diagnostic guidelines as applied to the particular service user.

Measurement of clinical utility

On the basis of earlier descriptions of the concept (Keeley, Reed, Roberts, Evans, Medina-Mora et al., 2016; Reed, 2010), the clinical utility of a classification construct or category for mental and behavioral disorders depends on its: (a) ease of communication (e.g., among practitioners, patients, families, administrators); (b) implementation characteristics in clinical practice, including goodness of fit (i.e., accuracy of description), ease of use and the time required to use it (i.e., feasibility); and (c) usefulness in selecting interventions and making clinical management decisions.

Accordingly, in the present study, the clinical utility of ICD-11 diagnostic guidelines was evaluated using a 4-point Likert scale to rate the different elements of these domains through a self-reported questionnaire applied to a particular patient. This scale was developed for the field studies designed to test the modifications proposed for ICD 11 (Keeley, Reed, Roberts, Evans, Medina-Mora et al., 2016; Reed, 2010) (see Table 7). Its factorial structure and internal consistency were evaluated prior to the main analyses regarding the clinicians' perception of the guidelines' clinical utility.

Statistical analyses

General characteristics of clinicians and patients were described using means and standard deviations for continuous variables and frequencies and percentages for

categorical variables. All the variables were compared between protocols (1 and 2), using independent sample *t*-tests or chi-square tests depending on the type of variables. Frequencies and percentages were also calculated to evaluate the general level of agreement (No agreement/Overall agreement) between interviewers and observers across all diagnostic groupings. Comparisons of frequencies of each diagnosis provided by the interviewer and observer were made using McNemar tests. Kappa values were calculated in order to summarize the level of diagnostic agreement between interviewers and observers.

Basic psychometric properties of the clinical utility measurement were obtained by calculating an exploratory factor analysis (using likelihood maximum extraction, Oblimin rotation and Kaiser-Meyer-Olkin measure of sampling adequacy (KMO), and a confirmatory model (IBM SPSS Amos 21) for factorial or construct validity, as well as total and subtotal Cronbach's alphas for internal consistency or reliability.

Lastly, in order to analyze clinical utility information, the frequencies and percentages of each item were described for both interviewers and observers. Total means were compared between interviewers and observers using a *t*-test for independent samples. The significance level for all tests was established at $p = .05$.

Results

A total sample of 23 clinicians accredited to make diagnosis in Mexico (17 psychiatrists and six fourth- or fifth-year psychiatry residents) evaluated 53 patients for Protocol 1 (with psychotic symptoms) and 100 patients for Protocol 2 (with mood, anxiety- or stress-related symptoms, without psychotic symptoms). Table 1 presents the basic clinician characteristics. No differences by gender, age, or professional experience were found between interviewers and observers. Participants' sociodemographic and clinical characteristics are presented in Table 2.

Clinician rater dyads for the evaluation of each participant were assigned on the basis of a systematic sampling procedure using a list of clinicians available each day and taking into account their most recent role as observer or interviewer in order to maximize the variability of dyads and roles. Accordingly, the percentage of repeated dyads was less than half the total number of dyads.

Diagnostic agreement with and without ICD-11 guidelines

Table 3 presents the Kappa's coefficients for the diagnostic guidelines of each ICD-11 diagnostic group.

Clinical utility of ICD-11 diagnostic guidelines

The Scale of Clinical Utility of the ICD-11 Mental and Behavioural diagnostic guidelines was first evaluated in terms of its construct validity (factorial validity) and reliability (internal consistency). Table 4 presents the results of the exploratory factorial analysis of the scale, as well as internal consistency coefficients for the total and subtotal scores.

Table 1 Demographics and years of experience between interviewers and observers.

	Interviewer		Observer		Comparison
	Mean	SD	Mean	SD	
Age	37.6	9.0	35.5	7.5	$t(138) = 1.33$ $p = .184$
Professional experience (years)	6.6	7.6	6.7	6.9	$t(304) = -1.18$; $p = .906$
Sex	<i>n</i>	%	<i>n</i>	%	
Male	77	50.3	72	47.1	$\chi^2 (1) = 0.20$; $p = .647$
Female	76	49.7	81	52.9	

Table 2 Demographics: Patients in protocols 1 and 2.

	Protocol 1 With psychotic symptoms <i>n</i> =53		Protocol 2 Mood/anxious/stress-related <i>n</i> =100		Comparison Protocol 1 vs. Protocol 2
	Media	SD	Media	SD	
Age	36.7	11.9	38.2	13.6	$t(151) = -0.67$; $p = .500$
Sex	<i>n</i>	%	<i>n</i>	%	
Male	27	50.9	19	19.0	$\chi^2 (1) = 15.32$; $p < .001$
Female	26	49.1	81	81.0	
Civil Status					
Single/separated/divorced	48	90.6	61	61.0	$\chi^2 (1) = 13.37$; $p < .001$
Married/Cohabiting	5	9.4	39	39.0	
Work Status Employee	11	20.8	36	36.0	$\chi^2 (2) = 5.77$; $p = .059$
Unemployed/Retired	36	67.9	48	48.0	
Student	6	11.3	16	16.0	

Table 3 Agreement between interviewers and observers.

Observer		Interviewer				Kappa
		Yes		No		
		f	%	f	%	
Schizophrenia and Other Primary Psychotic Disorders	Yes	42	85.7	7	14.3	.83 [*]
	No	4	3.8	100	96.2	
Mood disorders	Yes	89	87.3	13	12.7	.60 [*]
	No	14	27.5	37	72.5	
Anxiety- and Fear-Related Disorders	Yes	27	61.4	17	38.6	.43 [*]
	No	19	17.4	90	82.6	
Disorders Specifically associated with Stress	Yes	43	89.6	5	10.4	.77 [*]
	No	10	9.5	95	90.5	
Other disorders for which ICD-11 diagnostic guidelines had not been provided	Yes	21	50	21	50	.35 [*]
	No	17	15.3	94	84.7	

* $p \leq .001$

Two factors with eigen values above 1 together account for 60% of the variance. These factors involved the same general type of items. Factor one grouped together items regarding the clinical utility of the guidelines for case identification and management, while factor two included items concerning the evaluation of implementation

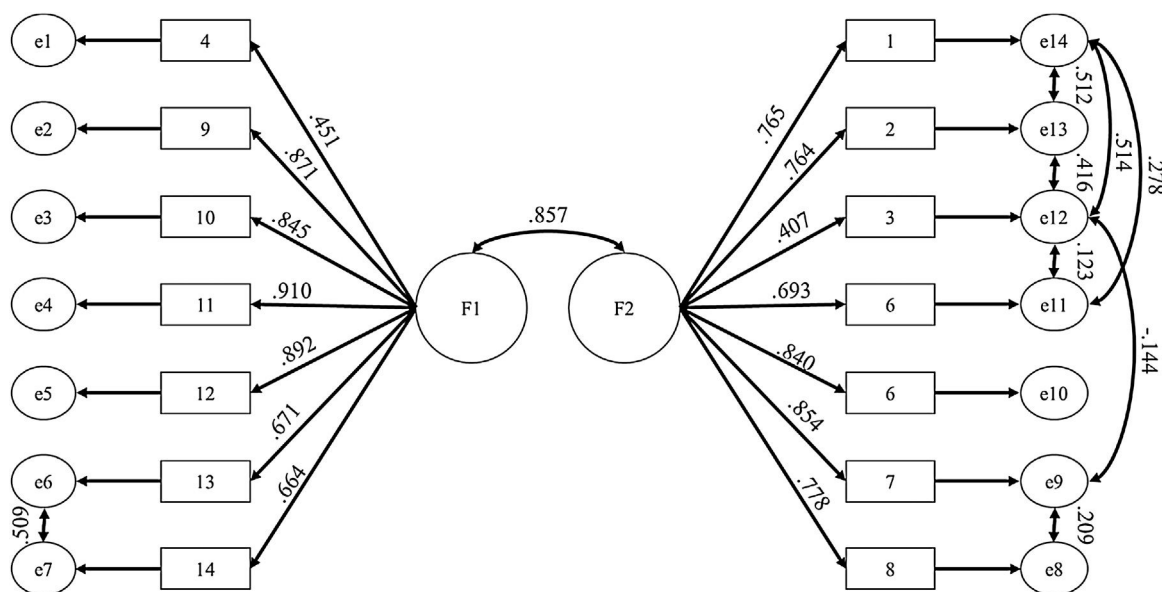
characteristics. Cronbach's alphas were over .90 for the total and subtotal scores.

Confirmatory model is presented in Figure 1, showing a good adjustment of the factor structure resulted from the exploratory analysis ($\chi^2 = 95.69$, $df = 68$, $p = .015$, $GFI = .956$, $RMR = .01$, $CFI = .991$, $RMSEA = .038$, $CI90\% = .017-.054$).

Table 4 Scale of Clinical Utility: Factorial validity and internal consistency.

	Identification & management	Implementation characteristics
4. Level of detail	.34	
9. Selection of treatment	.85	
10. Prognosis	.81	
11. Communicate	.79	
12. Educate	.90	
13. Qualifiers to select a treatment	.75	
14. Qualifiers and prognosis	.72	
1. Ease of use		-.99
2. Goodness of fit or accuracy		-.83
3. Clear and understandable		-.90
5. Difficult to assess		-.67
6. Amount of time		-.37
7. Boundary with normality		-.42
8. Boundary between disorders		-.47
Percentage of Variance	54.26	5.78
Cronbach alpha	.90	.901
Cronbach's alpha total scale	.93	

Note: $n = 306$ (observers & interviewers (real $n = 287$ with 19 missing values); Maximum extraction: likelihood; Oblimin rotation; Total percentage of variance explained = 60.04; Kaiser-Meyer-Olkin measure of sampling adequacy (KMO) = 0.938.

**Figure 1** Scale of Clinical Utility: Confirmatory model.

Note: $n = 287$ with the whole sample of interviewers and observers; $\chi^2 = 95.69$, $df = 68$, $p = .015$, $GFI = .956$, $RMR = .01$, $CFI = .991$, $RMSEA = .038$, $CI90\% = .017-.054$.

According to this scale, a high proportion of clinicians considered that all the diagnostic guidelines studied are quite or extremely useful (Table 5).

In general terms, the more frequent answer option was, by far, the one referring to a good clinical utility (i.e., quite easy to use, quite easy to apply, quite useful, etc.), following the one related to a very good clinical utility (i.e., extremely easy to use, extremely easy to apply, extremely useful, etc.). When adding the frequency of both answer options, the clinical utility of the ICD-11 guidelines under

study, given their implementation characteristics (ease of use, goodness of fit, clarity, amount of time required, etc.), were good or very good for more than 85% of the clinicians; ranging between 85.6% (for the description of the boundary between the patient's disorder and other disorders) to 91.5% (concerning their ease of use and clarity). Consistently, the clinical utility of the guidelines for identification and management of cases (including their utility to communicate with and educate patients and family) was rated as good or very good by more than 80% of the clinicians; ranging between

Table 5 Clinical utility measure: Items and frequencies of responses by clinicians with respect to all patients (N=153).

Items	Answer options	Interviewer n = 153		Observer n = 153		Total N = 306	
		f	%	f	%	f	%
1. Please rate the overall EASE OF USE of the diagnostic guidelines with respect to this patient.	Not at all easy to use	1	0.7	-	-	1	0.3
	Somewhat easy to use	10	6.5	15	9.8	25	8.2
	Quite easy to use	107	69.9	106	69.3	213	69.6
	Extremely easy to use	35	22.9	32	20.9	67	21.9
2. Please rate the overall GOODNESS OF FIT or ACCURACY of the diagnostic guidelines...	Not at all accurate	1	0.7	1	0.7	2	0.7
	Somewhat accurate	18	11.8	20	13.1	38	12.4
	Quite accurate	106	69.3	102	66.7	208	68.0
	Extremely accurate	28	18.3	30	19.6	58	19.0
3. Please rate the extent to which the diagnostic guidelines were CLEAR AND UNDERSTANDABLE...	Not at all / somewhat clear	10	6.5	17	11.1	27	8.8
	Quite clear and...	108	70.6	105	68.6	213	69.6
	Extremely clear and...	35	22.9	31	20.3	66	21.6
	Insufficient detail and...	14	9.2	19	12.4	33	10.8
4. Which of the following statements best describes your evaluation of the LEVEL OF DETAIL AND SPECIFICITY...	About the right amount of...	137	89.5	127	83.0	264	86.3
	Too much detail and...	2	1.3	7	4.6	9	2.9
	Very difficult to apply	1	0.7	1	0.7	2	0.7
	Somewhat difficult to apply	17	11.1	23	15.0	40	13.1
5. Please rate to the extent to which the guidelines imposed requirements that were DIFFICULT TO ASSESS...	Quite easy to apply	110	71.9	110	71.9	220	71.9
	Extremely easy to apply	25	16.3	19	12.4	44	14.4
	Longer than my usual clinical practice	15	9.8	19	12.4	34	11.1
	About the same as my usual	95	62.1	89	58.2	184	60.1
6. How would you describe the AMOUNT OF TIME that it took you to apply all of the essential features...	Shorter than my usual	43	28.1	45	29.4	88	28.8
	Not at all useful	1	0.7	1	0.7	2	0.7
	Somewhat useful	14	9.2	17	11.1	31	10.1
	Quite useful	114	74.5	106	69.3	220	71.9
7. Please rate the extent to which the description of the BOUNDARY BETWEEN DISORDER AND NORMALITY...	Extremely useful	24	15.7	29	19.0	53	17.3

Table 5 (Continued)

Items	Answer options	Interviewer <i>n</i> = 153		Observer <i>n</i> = 153		Total <i>N</i> = 306	
		<i>f</i>	%	<i>f</i>	%	<i>f</i>	%
8. Please rate the extent to which the description of the BOUNDARY BETWEEN THIS PATIENT'S DISORDER A OTHER DISORDERS...	Not at all useful	-	-	2	1.3	2	0.7
	Somewhat useful	20	13.1	22	14.4	42	13.7
	Quite useful	106	69.3	104	68.0	210	68.6
	Extremely useful	27	17.6	25	16.3	52	17.0
9. How useful would the diagnostic guidelines be in helping you to SELECT A TREATMENT for this patient?	Not at all useful	-	-	4	2.6	4	1.3
	Somewhat useful	15	9.8	21	13.7	36	11.8
	Quite useful	105	68.6	95	62.1	200	65.4
	Extremely useful	33	21.6	33	21.6	66	21.6
10. How useful would the diagnostic guidelines be in helping you to assess this patient's PROGNOSIS?	Not at all useful	-	-	3	2.0	3	1.0
	Somewhat useful	21	13.7	24	15.7	45	14.7
	Quite useful	101	66.0	93	60.8	194	63.4
	Extremely useful	31	20.3	33	21.6	64	20.9
11. How useful would the diagnostic guidelines be in helping you to COMMUNICATE about this patient...	Not at all useful	1	0.7	24	15.7	3	1.0
	Somewhat useful	13	8.5	129	84.3	35	11.4
	Quite useful	102	66.7	-	-	195	63.7
	Extremely useful	37	24.2	-	-	73	23.9
12. How useful would the diagnostic guidelines be in helping you to EDUCATE this patient and/or family...	Not at all useful	-	-	2	1.3	2	0.7
	Somewhat useful	19	12.4	22	14.4	41	13.4
	Quite useful	97	63.4	96	62.7	193	63.1
	Extremely useful	37	24.2	33	21.6	70	22.9
13. How useful would the QUALIFIERS be in helping you to SELECT A TREATMENT for this patient?	Not at all useful	1	0.7	-	-	1	0.3
	Somewhat useful	12	8.5	19	12.9	31	10.7
	Quite useful	57	40.1	67	45.6	124	42.9
	Extremely useful	72	50.7	61	41.5	133	46.0
14. How useful would the QUALIFIERS be in helping you to determine this patient's PROGNOSIS?	Not at all useful	-	-	1	0.7	1	0.3
	Somewhat useful	17	11.9	24	16.4	41	14.2
	Quite useful	68	47.6	57	39.0	125	43.3
	Extremely useful	58	40.6	64	43.8	122	42.2
Total clinical utility *	Mean	SD	Mean	SD	Mean	SD	
	28.4	6.0	26.7	6.1	28.2	6.2	

Note. * *t* (152) = 2.57, *p* = 0.11

81.3% (for guidelines' utility to assess patient's prognosis) to 88.9% (for qualifiers as helpful to select a treatment for the patient).

Discussion

Reliable, clinically useful, and globally applicable diagnostic classification is an essential tool for reducing the treatment gap and the burden of disease attributable to common mental disorders in adulthood ([International Advisory Group for the Revision of ICD-10 Mental and Behavioural Disorders, 2011](#)). This is especially true in Latin American countries such as Mexico, where patients in need of care are not identified in a timely manner and only obtain treatment when their disorders are already very severe ([Borges et al., 2006](#); [Wang et al., 2007](#)), after having experienced a great deal of preventable suffering and disability.

Before discussing our results, several limitations of our study need to be considered. The sample is small, comprising a total of 153 patients independently evaluated by a pair of psychiatrists and medical doctors in training. Moreover, the data are drawn from a single institution oriented towards research, which also serves as a teaching hospital. The clinicians were psychiatrists and residents in training to become psychiatrists, who likely had high levels of training in comparison to the general population of clinicians. Despite these limitations, the results have significant implications for the implementation of the ICD-11 in Mexico and Latin-American countries.

Inter-rater reliability of ICD-11 diagnostic guidelines

According to Cohen's criteria ([Cohen, 1960](#)), diagnostic agreement between raters using ICD-11 guidelines can be rated as strong for Schizophrenia and Other Primary Psychotic Disorders, moderate for mood and stress-related disorders, and weak -although acceptable- for anxiety and fear-related disorders. However, consistent with [McHugh \(2012\)](#), "*Cohen's suggested interpretation may be too lenient for health-related studies because it implies that a score as low as 0.41 might be acceptable*" (pp. 276), and being strict, a kappa below to 0.60 as in the case of diagnoses for anxiety and fear-related disorders, indicates inadequate agreement among the raters.

This might be explained in part given such group of disorders was less common in the sample. Consequently, specificity was high for all the diagnostic groups under study while sensitivity was lower for anxiety and fear-related disorders (as well as other diagnoses). Another plausible explanation could be related with the high comorbidity of anxiety disorders with the other diagnosis under study. Thus, it is possible that, being such a common manifestation, hinders the diagnostic separation even though clinically it is more accessible for expert clinicians.

Still, although we did not provide any guidance on how the interview was to be conducted, and the majority of the cases presents with a clinically significant severity and comorbidity (given they were recruited in a specialized institution), observed kappa indexes were similar to those achieved using more complex and time-consuming

instruments (such as structured or semi-structured clinical interviews) ([Pies, 2007](#)). And even though our results are not comparable to DSM-5 reliability studies, which used a different methodology, they challenge the assumption that a less rigid diagnostic guidelines are inherently less reliable ([Craddock & Mynors-Wallis, 2014](#)), probably because in their attempts to communicate the essence of the disorder, they are more similar to how clinicians think.

Clinical utility of ICD-11 diagnostic guidelines

The present study also provides information on the perception of clinicians regarding the clinical utility of the diagnostic guidelines evaluated. This is important because of the emphasis on increasing the clinical utility ([Keeley, Reed, Roberts, Evans, Robles et al., 2016](#); [Reed, 2010](#)) of the classification as a whole in order to provide a tool that will help reduce the global burden of disease through early identification and the treatment of health conditions.

Regarding the diagnostic guidelines for psychotic, mood, anxiety and stress-related disorders proposed for ICD-11, we can infer from our results that Mexican clinicians with extensive experience of attending psychiatric patients consider that they are of value in terms of their implementing characteristics (mainly regarding their ease of use and clarity) as well as for the identification and management of patients, specially their qualifiers to select a specific treatment. This important finding (given that the ultimate goal of a clinical useful classification is to help in the decision of a proper case management) seems to be in line with several WG's proposals, including a different system of qualifiers for Schizophrenia and other Psychotic Disorders, which considers the evaluation of the level of cognitive impairment that may indicate the need for cognitive remediation interventions.

Additionally, concerning the classification of depressive disorders, one of the common mental disorders responsible for a large burden of disease in Mexico, Latin America and globally ([Medina-Mora et al., 2007](#); [World Health Organization WHO, 2017](#)), although ICD-11 classification was not been substantially modified, the proposed diagnostic guidelines include new severity qualifiers that were expected to improve their clinical utility ([Chakrabarti, Berlanga, & Njenga, 2012](#)) especially regarding treatment selection, which might vary considerably from a mild to a severe case. However, there are some space for additional improvements, mainly in terms of the guidelines' utility to assess patient's prognosis, which could require, in many cases, not just a systematic effort to include the information needed to do so, but the generation of such scientific data by psychiatric entity. An additional contribution of this study is the psychometric evaluation of the Scale to Measure Clinical Utility ([Keeley, Reed, Roberts, Evans, Medina-Mora et al., 2016](#); [Reed, 2010](#)) in a reliable, valid manner for future studies in the field.

According to our results, the ICD-11 would appear to constitute a reliable, clinically useful diagnostic system, at least as regards clinician consistency when the guidelines are used to identify mental disorders that account for the greatest proportion of years lived with disability, and for which there is a considerable treatment gap in both developed

and developing countries (Pan American Health Organization PAHO, 2013; Wang et al., 2007).

Acknowledgements

The study was supported by the National Council of Science and Technology (CONACyT) of Mexico, Project number 234473 and by the Instituto Nacional de Psiquiatría Ramón de la Fuente Muñiz, Mexico. The authors wish to acknowledge the important work by research assistants Omar Hernández, Alejandra Gonzalez, Carolina Muñoz, Lucia Munch Tania Real, and all the residents and psychiatrists who participated in this study.

References

- Benjet, C., Borges, G., & Medina-Mora, M. E. (2010). Chronic childhood adversity and onset of psychopathology during three life stages: Childhood, adolescence and adulthood. *Journal of Psychiatric Research*, 44, 732–740.
- Borges, G., Medina-Mora, M. E., Wang, P. S., Lara, C., Berglund, P., & Walters, E. (2006). Treatment and adequacy of treatment of mental disorders among respondents to the Mexico National Comorbidity Survey. *American Journal of Psychiatry*, 163, 1371–1378.
- Bromet, E. J., Atwoli, L., Kawakami, N., Navarro-Mateu, F., Piotrowski, P., King, A. J., Aguilar-Gaxiola, S., Alonso, J., Bunting, B., Demyttenaere, K., Florescu, S., de Girolamo, G., Gluzman, S., Haro, J. M., de Jonge, P., Karam, E. G., Lee, S., Kovess-Masfety, V., Medina-Mora, M. E., Mneimneh, Z., Pennell, B. E., Posada-Villa, J., Salmerón, D., Takeshima, T., & Kessler, R. C. (2017). Post-traumatic stress disorder associated with natural and human-made disasters in the World Mental Health Surveys. *Psychological Medicine*, 47, 227–241.
- Chakrabarti, S., Berlanga, C., & Njenga, F. (2012). Cultural issues in the classification and diagnosis of mood and anxiety disorders. *World Psychiatry*, 11(Suppl. 1), 26–30.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46.
- Craddock, N., & Mynors-Wallis, L. (2014). Psychiatric diagnosis: Impersonal, imperfect and important. *The British Journal of Psychiatry*, 204, 93–95.
- First, M. B., Reed, G. M., Saxena, S., & Hyman, S. E. (2015). The development of the ICD-11 clinical descriptions and diagnostic guidelines for mental and behavioral disorders. *World Psychiatry*, 14, 82–90. <http://dx.doi.org/10.1002/wps.20189>
- 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. doi: 10.1002/j.2051-5545.2011.tb00022.x.
- Keeley, J. W., Reed, G. M., Roberts, M. C., Evans, S. C., Medina-Mora, M. E., Robles, R., Rebello, T., Sharan, P., Gureje, O., First, M. B., Andrews, H. F., Ayuso-Mateos, J. L., Gaebel, W., Zielasek, J., & Saxena, S. (2016). Developing a science of clinical utility in diagnostic classification systems: Field study strategies for ICD-11 mental and behavioral disorders. *American Psychologist*, 71, 3–16. <http://dx.doi.org/10.1037/a0039972>
- Keeley, J. W., Reed, G. M., Roberts, M. C., Evans, S. C., Robles, R., Matsumoto, C., Brewin, C. R., Cloitre, M., Perkonig, A., Rousseau, C., Gureje, O., Lovell, A. M., Sharan, P., & Maercker, A. (2016). Disorders specifically associated with stress: A case-controlled field study for ICD-11 Mental and Behavioural Disorders. *International Journal of Clinical and Health Psychology*, 16, 109–127. <http://dx.doi.org/10.1016/j.ijchp.2015.09.002>
- Kessler, R., Angermeyer, M., Anthony, J., de Graaf, R., Demyttenaere, K., Gasquet, I., de Girolamo, G., Gluzman, S., Gureje, O., Haro, J. M., Kawakami, N., Karam, A. N., Levinson, D., Medina-Mora, M. E., Browne, M. A. O., Posada-Villa, J., Stein, D. J., Tsang, C. H. A., Aguilar-Gaxiola, S., Alonso, J., Lee, S., Heeringa, S., Pennell, B. E., Berglund, P., Gruber, M. J., Petukhova, M., Chatterji, S., & Üstün, T. B. (2007). Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Surveys. *World Psychiatry*, 6, 168–176.
- Liu, H., Petukhova, M., Sampson, N. A., Aguilar-Gaxiola, S., Alonso, J., Andrade, L. H., Bromet, E. J., de Girolamo, G., Haro, J. M., Hinkov, H., Kawakami, N., Koenen, K. C., Kovess-Masfety, V., Lee, S., Medina-Mora, M. E., Navarro-Mateu, F., O'Neill, S., Piazza, M., Posada-Villa, J., Shahly, V., Stein, D. J., Ten Have, M., Torres, Y., Gureje, O., Zaslavsky, A. M., Kessler, R., & World Mental Health Survey Collaborators. (2017). Association of DSM-IV posttraumatic Stress Disorder With Traumatic Experience Type and History in the World Health Organization World Mental Health Surveys. *JAMA Psychiatry*, 74, 270–281. <http://dx.doi.org/10.1001/jamapsychiatry.2016.3783>
- McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica: Biochemia Medica*, 22, 276–282.
- Medina-Mora, M. E., Borges, G., Benjet, C., Lara, C., & Berglund, P. (2007). Psychiatric disorders in Mexico: lifetime prevalence in a nationally representative sample. *British Journal of Psychiatry*, 190, 521–528.
- Navarro, F., Alonso, J., Lim, C. C. W., Saha, S., Aguilar-Gaxiola, S., Al-Hamzawi, A., Andrade, L. H., Bromet, E. J., Bruffaerts, R., Chatterji, S., Degenhardt, L., de Girolamo, G., de Jonge, P., Fayyad, J., Florescu, S., Gureje, O., Haro, J. M., Hu, C., Karam, E. G., Kovess-Masfety, V., Lee, S., Medina-Mora, M. E., Ojagbemi, A., Pennell, B. E., Piazza, M., Posada-Villa, J., Scott, K. M., Stagnaro, J. C., Xavier, M., Kendeler, K. S., Kessler, R., McGrath, J. J., & WHO World Mental Health Survey Collaborators. (2017). The association between psychotic experiences and disability: Results from the WHO World Mental Health Surveys. *Acta Psychiatrica Scandinavica*, 136, 74–84. <http://dx.doi.org/10.1111/acps.12749>. Epub 2017 May 25
- Pan American Health Organization (2013). Mental Health in the Americas, 2013. Retrieved from <https://www.paho.org/saluden-las-americas-2017/?tag=mental-health>.
- Pies, R. (2007). How "objective" are psychiatric diagnoses?: (guess again). *Psychiatry (Edgmont)*, 4, 18.
- Reed, G. M. (2010). Toward ICD-11: Improving the clinical utility of WHO's international classification of mental disorders. *Professional Psychology: Research and Practice*, 41, 457–464. <http://dx.doi.org/10.1037/a0021701>
- Reed, G. M., Rebello, T. J., Pike, K. M., Medina-Mora, M. E., Gureje, O., Zhao, M., Dai, Y., Roberts, M. C., Maruta, T., Matsumoto, C., Krasnov, V. N., Kulygina, M., Lovell, A. M., Stona, A. C., Sharan, P., Robles, R., Gaebel, W., Zielasek, J., Khoury, B., Mari, J. J., Ayuso-Mateos, J. L., Evans, S. C., Kogan, C. S., & Saxena, S. (2015). WHO's Global Clinical Practice Network for Mental Health. *Lancet Psychiatry*, 2, 379–380. [http://dx.doi.org/10.1016/S2215-0366\(15\)00183-2](http://dx.doi.org/10.1016/S2215-0366(15)00183-2)
- Reed, G. M., Sharan, P., Rebello, T., Keeley, J., Medina-Mora, M. E., Gureje, O., Ayuso-Mateos, J. L., Kanba, S. H., Khoury, B., Kogan, C., Krasnov, V., Maj, M., Mari, J., Stein, D., Zhao, M., Akiyama, T., Andrews, H., Asevedo, E., Cheour, M., Domínguez-Martínez, T., El-Khoury, J., Fiorillo, A. Q., Grenier, J., Gupta, N., Kola, L., Kulygina, M., Leal-Leturia, I., Luciano, M., Lusu, B., Martínez-López, N., Matsumoto, H., Onofa, L., Paterniti, S., Purnima, Sh., Robles, R., Sahu, M., Sibeko, G., Zhong, N., First, M., Gaebel, W., Lovell, A., Maruta, T., Roberts, M., & Pike, K. (2018). The ICD-11 developmental field study of reliability of diagnoses of high-burden mental disorders: Results among adult patients in

- mental health settings of 13 countries. *World Psychiatry*, 17, 174–196.
- United Nations Office on Drugs and Crime UNODC. (2013). Global Study on Homicide 2013, Vienna, Austria. Retrieved from <http://www.unodc.org/gsh/>.
- Wang, P. S., Aguilar-Gaxiola, S., Alonso, J., Anger-meyer, M. C., Borges, G., Bromet, E. J., Bruffaerts, R., de Girolamo, G., de Graaf, R., Gureje, O., Haro, J. M., Karam, E. G., Kessler, R. C., Kovess, V., Lane, M. C., Lee, S., Levinson, D., Ono, Y., Petukhova, M., Posada-Villa, J., Seedat, S., & Wells, J. E. (2007). Use of Mental health services for anxiety, mood and substance disorders in 17 countries in the WHO World Mental Health Surveys. *The Lancet*, 370, 871-850.
- World Health Organization WHO (2017). *Depression and Other Common Mental Disorders: Global Health Estimates*. Geneva: World Health Organization.