Received: 5 May 2020 | Revised: 24 December 2020 | Accepted: 21 January 2021

DOI: 10.1002/jclp.23125

RESEARCH ARTICLE

WILEY

Taxonomy and utility in the diagnostic classification of mental disorders

Spencer C. Evans[1] | Michael C. Roberts[2] | Jessy Guler[2] | Jared W. Keeley[3] | Geoffrey M. Reed[4,5]

[1]Department of Psychology, University of Miami, Coral Gables, Florida, USA

[2]Clinical Child Psychology Program, University of Kansas, Lawrence, Kansas, USA

[3]Department of Psychology, Virginia Commonwealth University, Richmond, Virginia, USA

[4]Department of Psychiatry, Columbia University Vagelos College of Physicians and Surgeons, New York, New York, USA

[5]Department of Mental Health and Substance Abuse, World Health Organization, Geneva, Switzerland

Correspondence

Spencer C. Evans, Department of Psychology, University of Miami, Coral Gables, FL, USA.
Email: sevans@miami.edu

Funding information

National Institute of Mental Health, Grant/Award Number: Loan Repayment Program; AIM for Mental Health, Grant/Award Number: Clinical Science Fellowship

Abstract

Objective: One strategy for improving the clinical utility of mental health diagnostic systems is to better align them with how clinicians conceptualize psychopathology in practice. This approach was used in International Classification of Diseases 11th Revision (ICD-11) development, but its underlying assumption—a link between taxonomic "fit" and clinical utility—remains untested.

Methods: Using data from global mental health clinician samples (combined $N = 5404$), we investigated the association between taxonomic fit and clinical utility in mental disorder categories.

Results: The overall association between fit and utility was positive ($r = 0.19$) but statistically not different from zero (95% confidence interval [CI]: −0.06, 0.43) in this small sample ($N = 39$ ICD/DSM categories). However, a positive association became clear after correcting for outliers ($r = 0.34$ [0.05, 0.58] or higher). Further insights were apparent for specific diagnoses given their locations in the scatterplot.

Conclusions: Results suggest a positive link between taxonomic fit and clinical utility in mental disorder diagnoses, highlighting future research directions.

KEYWORDS

classification systems, clinical utility, diagnosis, *International Classification of Diseases 11th Revision*, taxonomy

1 | INTRODUCTION

In the revision of diagnostic classification systems of mental disorders, a central goal is to improve clinical utility. The rationale for this goal is that improvements in clinical utility may facilitate numerous "downstream" benefits, including improvements in diagnosis, clinical care, statistical reporting, and allocation of health resources. Among several potential avenues for improving clinical utility, one strategy—adopted in the development of the Eleventh Revision of the International Classification of Diseases and Related Health Problems (ICD-11) chapter on Mental, Behavioral, and Neurodevelopmental Disorders—is to adjust the taxonomies and diagnostic entities to better "fit" with clinicians' conceptualizations of mental disorders (Reed, 2010; Reed et al., 2013; Roberts et al., 2012). The premise is that diagnostic structures and entities that "fit" together in a cohesive, coherent manner may also be easier for clinicians to use in practice. While this is an interesting idea, with potentially profound implications, it has never been directly investigated. Accordingly, the present study explores the association between taxonomic fit and clinical utility in mental disorder classification.

1.1 | Goals of mental health classification

Diagnostic classification systems provide guidance for most efforts to understand, assess, and treat mental illness. These systems not only specify the names, descriptions, and diagnostic requirements for mental disorder diagnoses, they also organize the disorders into diagnostic categories, groups, and ultimately into a complex multilevel taxonomy. Globally, there are two dominant diagnostic classification systems used in regular clinical practice. One is the *Diagnostic and Statistical Manual of Mental Disorders* (DSM), from the American Psychiatric Association (APA), currently in its 5th edition (DSM-5; APA, 2013). The other is the *ICD* chapter on Mental, Behavioral, and Neurodevelopmental Disorders, developed by the World Health Organization (2020) and used by the large majority of mental health professionals worldwide (Evans et al., 2013; First et al., 2018; Reed et al., 2011). A comparison of these two systems and other frameworks is beyond the scope of this paper (for a review, see Clark et al., 2017). But a recent analysis by First et al. revealed that 80.4% of 102 diagnostic entities examined were essentially identical or showed only minor differences between ICD-11 and DSM-5. Accordingly, while the present research was born out of the ICD-11 development, the more important point is that the DSM and ICD are substantially similar—different systems with a largely common lexicon for the classification and diagnosis of mental disorders.

The development of the ICD-11 (World Health Organization, 2020)—a recent revision of the ICD-10 (World Health Organization 1992a, 1992b)—was an international and empirically driven undertaking that took more than a decade, with implications for mental health practice and research around the globe (Reed et al., 2019; Keeley et al., 2016). The ICD-11 was adopted by the 72nd World Health Assembly in 2019, with a plan to put it into effect starting in 2022 (World Health Organization, 2019). The revision of any diagnostic system is a substantial undertaking. Recent revisions have begun by establishing a research agenda with defined goals and priorities for the new editions (First et al., 2015; International Advisory Group for the Revision of ICD-10 Mental and Behavioral Disorders, 2011; Keeley et al., 2016; Kupfer et al., 2002, 2008; Reed, 2010). Fundamentally, these agendas have two priorities regarding mental health diagnosis: validity and utility (Kendall & Jablensky, 2003). Whereas modern thinking about *validity* often involves the extent to which a diagnosis represents a distinct "disease entity" with natural boundaries (e.g., genetic markers, pathophysiology, neural substrates), *utility* essentially refers to whether a diagnostic construct is useful in clinical research and practice. It has long been hoped that diagnostic validity could be improved by incorporating recent biomarker findings from fields such as genetics and neuroscience. However, validity is an elusive goal, and research advances in these areas have not been sufficient to effect major changes to ICD or DSM (Hyman, 2010; Insel et al., 2010; Reed, 2010).

Accordingly, recent revision efforts have renewed their focus on the priority of clinical utility. While there are several definitions of clinical utility (First et al., 2004; Mullins-Sweatt & Widiger, 2009), Reed's (2010) definition is

useful for the present study because it was operationalized for research:

> The clinical utility of a classification construct or category for mental and behavioral disorders depends on: a) its value in *communicating* (e.g., among practitioners, patients, families, administrators); b) its *implementation characteristics* in clinical practice, including its goodness of fit (i.e., accuracy of description), its ease of use, and the time required to use it (i.e., feasibility); and c) its usefulness in *selecting interventions* and in making *clinical management* decisions. (p. 461)

Although scholarship on clinical utility is a recent development, the goal of clinical utility is as old as the very enterprise of describing mental disorders (Engstrom & Kendler, 2015). The texts of ICD and DSM attest to the constancy of this goal. Like its predecessors, DSM-5 states that it was "designed first and foremost to be a useful guide to clinical practice" (American Psychiatric Association, 2013, p. xli). Similarly, ICD-10-CM was published specifically for "general clinical, educational and service use" (World Health Organization, 1992b, p. 8). Over the years, many have argued that revisions should be made to still further improve clinical utility (e.g., First et al., 2018; Flanagan & Blashfield, 2007b, 2010; Keeley et al., 2016; Mullins-Sweatt & Widiger, 2009; Reed et al., 2013, 2018). Indeed, the revision processes for ICD and DSM involve a great deal of attention toward assessing and improving clinical utility across diagnostic classification categories (Clark et al., 2017; International Advisory Group for the Revision of ICD-10 Mental and Behavioral Disorders, 2011; Keeley et al., 2016; Regier et al., 2009).

1.2 | The empirical basis for taxonomy and utility

Despite this attention toward improving clinical utility in mental health classification, empirical knowledge of the mechanisms underlying clinical utility remains limited. Research in this area has tended to focus on clinical utility as an *outcome* in the evaluation of various specific categories. This knowledge about the usefulness of diagnostic categories would be greatly enhanced by learning about *why* they are useful. Here, research is only beginning to scratch the surface. Relevant evidence generally falls into three categories: (a) field trials, which examine the ease of use, perceived validity, and reliability of diagnostic systems (e.g., Evans et al., 2020; Keeley et al., 2016; Reed et al., 2018; Regier et al., 2013; Sartorius et al., 1993; Shaffer, 1996); (b) general surveys of clinicians, which examine variables such as usage, satisfaction, problems, and areas for improvement of diagnostic systems (e.g., Evans et al., 2013; First et al., 2018; Mellsop et al., 2007; Reed et al., 2011; Suzuki et al., 2010; Zielasek et al., 2010); and (c) studies of clinicians' cognition, including their working taxonomies of mental disorders (Flanagan & Blashfield, 2006, 2007a; Flanagan et al., 2008, 2012; Reed et al., 2013; Roberts et al., 2012).

These methodologies inform the assessment and improvement of clinical utility, but important questions persist regarding predictors of clinical utility. Field trials assess whether some improvement has been made in utility of a revised diagnostic system relative to a prior or alternative version, but they do not explain the mechanisms that are helping or hindering that improvement. Practitioner surveys are useful in that they give voice to the clinical users of diagnostic systems to indicate where and how they should be improved; yet, surveys essentially aim to identify respondents' patterns of perspectives and experiences, and these patterns may not converge on clear, actionable results. Studies of clinicians' cognitive views and processes are particularly promising. However, before the ICD-11 research, clinician cognition studies have tended to use small samples of clinicians available for in-person assessments, which limits the potential for reliable and quantifiable improvements to the utility of disorder classification systems on a global scale.

Steps toward global applicability have been taken in WHO field studies (Reed et al., 2013; Roberts et al., 2012), which were inspired by foundational work on clinicians' cognitive taxonomies. In particular, Flanagan et al. (2007a, 2007b, 2010) employed methods from anthropology and cognitive psychology to investigate clinicians' folk taxonomies of mental disorders. In these studies, clinician participants were given 67 cards, each containing the

label of a diagnostic category from the DSM-IV. They were then asked to sort them into groups that made sense based on their clinical similarity. Cluster analyses revealed taxonomies of mental disorders that were relatively consistent across clinicians. These taxonomies were simpler than and different from the DSM, although moderately correlated with it. Flanagan and Blashfield (2007b, 2010) later argued that the clinical utility of the DSM and ICD could be improved by revising them to more closely align with the way clinicians conceptualize and use mental disorder diagnoses in practice. Subsequently, WHO adopted a similar approach in two of their formative global research investigations for ICD-11 (Reed et al., 2013; Roberts et al., 2012). These studies employed paired-comparison and card-sorting methods to examine how clinicians view relationships among mental disorders. The findings informed the architecture of the ICD-11 for optimal clinical utility.

1.3 | The present study

In summary, as noted by Reed et al. (2013): "It is reasonable to assume—and a testable hypothesis—that the more compatible the organization of the classification is with clinicians' implicit structures of mental and behavioral disorders categories, the easier and more intuitive it will be for them to use the system" (p. 1209). The present study tests one version of this "testable hypothesis" by examining the extent to which clinical utility is associated with taxonomic fit in diagnostic categories. To our knowledge, this question has not been previously investigated; but such questions are amenable to research, with the key variables having already been operationalized in empirical studies (e.g., Evans et al., 2013; Reed et al., 2011, 2013; Roberts et al., 2012). Investigating the taxonomy-utility link could shed light on the principles underlying the ICD-11 revision while also offering much-needed guidance for improving clinical utility in future revisions of diagnostic classification systems.

Accordingly, the present study leverages data from two global clinician studies to examine the relationship between the clinical utility and the taxonomic fit of mental disorder categories. In the first study, Reed et al. (2011) conducted a global survey of psychiatrists who rated the clinical utility of numerous diagnostic categories. In the second study, Reed et al. (2013) investigated clinicians' working taxonomies of mental disorders according to their similarities and differences in clinical practice, yielding data about each category's taxonomic fit. Given certain characteristics of our two-sample approach (e.g., yielding a small sample size of $N = 39$ diagnostic categories), our focus was on the directionality and magnitude of effects, and formal tests of the null hypothesis were not appropriate. Still, we had some general hypotheses based on the literature reviewed above (e.g., Flanagan & Blashfield, 2007b, 2010; Reed, 2010). We expected to see a moderate, positive association between taxonomic fit and clinical utility. That is, diagnostic categories that "fit" better among other disorders were expected have higher clinical utility, compared to those with poorer fit.

2 | METHOD

2.1 | Participants

Data were extracted from the published reports of two studies (Reed et al., 2011, 2013). Full details of these investigations and their samples are provided in the original reports and summarized below. Both studies were carried out to inform the development of ICD-11 Mental and Behavioral Disorders. For the present purposes, the combined data represent a large sample (combined $N = 5404$) of mental health clinicians from globally representative cultural, professional, and linguistic backgrounds. Each study generated data from clinicians regarding their perceptions of specific mental disorder categories. All studies were approved by applicable institutional review boards.

2.1.1 | Global psychiatrists survey sample

Clinical utility ratings were drawn from a global survey of psychiatrists conducted in collaboration between the World Psychiatric Association and World Health Organization, in partnership with national societies of psychiatrists (Reed et al., 2011). Participants were 4887 psychiatrists (M_{Age} = 49.9 years; $M_{Experience}$ = 17.1 years; 62% male) from 44 countries in all global regions, participating in 19 languages. Country-level response rate averaged 34%, with higher response rates coming from lower-income countries. Nearly all respondents (97%) reported that they were currently seeing patients clinically, averaging about 26.9 direct patient contact hours per week. The large majority (79%) reported that they "often" or "almost always/always" used a diagnostic classification system; the dominant system was ICD-10 (used by 70%) followed by DSM-IV (23%) and various others (7% in total). Participating psychiatrists who reported using the ICD-10 were asked to select from a list of 44 diagnostic categories those that they used in day-to-day practice. Most selected 8–20. Then, they were asked questions about the categories they selected, drawing on their direct experience in using them clinically. These questions formed the basis for clinical utility data used here (see Section 2.2, below).

2.1.2 | Clinicians' natural taxonomies sample

Taxonomic fit data were drawn from an international study of the how psychologists and psychiatrists conceptualize the relationships among mental disorder categories (Reed et al., 2013). To be eligible, individuals were required to be a licensed or authorized mental health professional with 2+ years of posttraining experience, whose current activities included direct patient care for 10+ hours per week. The final sample included 517 mental health clinicians (M_{Age} = 42.3 years; $M_{Experience}$ = 11.9 years; $M_{Training}$ = 5.9 years; 57% male) from eight countries: Brazil China, India, Japan, Mexico, Nigeria, Spain, and the United States. The study was carried out in local languages of Chinese, English, Japanese, Portuguese, and Spanish. Professionally, 73% of participants were psychiatrists, 25% were psychologists, and 2% were in allied mental health professions (psychiatric nursing, counseling, and social work). Procedures involved meeting with an examiner for an experimental task involving sorting 60 cards, each labeled with a mental disorder category that was comprehensible to ICD and DSM users alike. Their sorting patterns formed the basis for our taxonomic fit data.

2.2 | Variables

2.2.1 | Clinical utility

Participants in the global psychiatrist survey sample (Reed et al., 2011) were asked to evaluate the clinical utility of the diagnostic categories they reported using in their regular clinical practice. Following a conceptual definition of clinical utility (Reed, 2010), these items asked participants to rate individual diagnostic categories in terms of their "ease of use" and their "goodness of fit." Ratings were originally made on a 4-point scale from 0 ("not at all easy to use in clinical practice" or "not at all accurate") to 3 ("extremely easy to use" or "extremely accurate"), then transformed to a scale from 0 (*low utility*) to 1 (*high utility*). For the present analysis, we averaged the 0–1 ease of use and goodness of fit ratings together to form a composite clinical utility rating for each applicable diagnosis.

2.2.2 | Taxonomic fit

Cohesion index values from the taxonomy sample (Reed et al., 2013) were used to represent the clinical utility of individual diagnostic categories. Participating clinicians were asked to sort mental disorder categories "into groups

that you think are clinically relevant in the assessment and management of people with mental and behavioral disorders." The resulting data were analyzed via cluster analysis to identify clinicians' general taxonomy of mental disorders. One byproduct of this analysis is a cohesion index estimated for each diagnostic category. Originally scaled from 0 (*high cohesiveness*) to 1 (*low cohesiveness*), we reverse-coded these data for the present analysis, from 0 (*low cohesiveness*) to 1 (*high cohesiveness*). That is, values closer to 0 indicate clinicians often disagreed about where the diagnostic category should be placed, whereas values closer to 1 indicated greater agreement.

2.3 | Analytic approach

Between the 44 diagnostic categories used in the survey study (Reed et al., 2011) and the 60 categories used in the taxonomy study (Reed et al., 2013), there were 39 diagnostic categories that both studies shared in common. The two quantitative variables (clinical utility, taxonomic fit) were available for all categories in the small sample ($N = 39$) and were standardized for analysis ($M = 0$, $SD = 1$). Given our small sample size and our focus on the directionality and magnitude of effects, null-hypothesis tests of statistical significance were not appropriate; bootstrapped 95% confidence intervals (95% CI) were used instead to quantify the uncertainty around the correlation estimates, in line with recent recommendations (Cumming, 2014; Halsey et al., 2015). Statistical inference proceeded in a circumspect manner, including visual inspection of the scatterplot with attention to possible outliers and the placement of individual data points (named disorder categories) relative to the axes and correlation reference lines.

3 | RESULTS

The names, clusters, utility ratings, and cohesiveness estimates of the 39 diagnostic categories are reported in Table 1. The scatterplot for the correlation between utility and cohesiveness is presented in Figure 1. The overall Pearson's correlation between cohesiveness and utility was positive but small in magnitude, $r = 0.19$ (95% CI = −0.06, 0.43), not statistically different from zero and accounting for relatively little of the overall variance in clinical utility ratings ($R^2 = 0.04$). The direction of the association provides preliminary support for the hypothesis that utility is modestly positively correlated with taxonomic fit.

Visual inspection of the scatterplot revealed insights that are not evident in the correlation coefficient. As shown in Figure 1, the positive relationship between clinical utility and taxonomic fit might be attenuated by a few outliers. To explore this possibility, we estimated the correlation again while dropping statistical outliers of both variables, with the outlier threshold operationalized at extreme (±2 *SD* from the mean) and moderate (±1.5 *SD* from the mean) levels. There were two extreme outliers: Tic Disorders and Anorexia Nervosa, which both had very low cohesiveness. With these two outliers filtered out, the correlation rose to $r = 0.34$ [0.05, 0.58], accounting for about 12% of the variance in clinical utility ($R^2 = 0.12$). Applying the moderate threshold for outlier detection (±1.5 *SD*), five additional categories were identified: Borderline Personality Disorder, Autistic Disorder, Social Phobia, Obsessive-Compulsive Disorder (OCD), and Adjustment Disorder. When these additional categories were removed, the correlation rose to $r = 0.58$ (0.34, 0.79), accounting for about one-third of the variance ($R^2 = 0.34$).

Lastly, the placement of individual data points in the scatterplot (Figure 1) offered further insights regarding the taxonomy-utility characteristics of specific categories. To systematize this analysis, we used reference lines for the observed correlation ($r = 0.19$), for a hypothetical perfect correlation ($r = 1$), for the major axes, and for ±1–2 *SD* from the mean. Thus, all data points can be viewed according to four quadrants. Points that are more extreme in the direction of their quadrants would be most characteristic of their quadrant, and therefore were given closer attention in interpretation. As shown, disorders with low utility and above-average cohesiveness (Figure 1, bottom right) include Autism Spectrum Disorder, Asperger's Syndrome, and Borderline Personality Disorder. On the

TABLE 1 Identifiers, clinical utility ratings, and taxonomic cohesiveness estimates for included mental disorder categories

Cluster[a]	Abbreviation	Diagnostic code and category[b]	Chapter[b]	Clinical utility[c]	Cohesiveness[a]
Anxiety OC&SR	Adj	F43.2 Adjustment disorders	F40–F48 Neurotic, stress-related & somatoform	0.62	0.26
Anxiety OC&SR	AnxDep	F41.2 Mixed anxiety and depressive disorder	F40–F48 Neurotic, stress-related & somatoform	0.60	0.31
Anxiety OC&SR	GAD	F41.1 Generalized anxiety disorder	F40–F48 Neurotic, stress-related & somatoform	0.67	0.49
Anxiety OC&SR	OCD	F42 Obsessive-compulsive disorder	F40–F48 Neurotic, stress-related & somatoform	0.73	0.38
Anxiety OC&SR	Panic	F41.0 Panic disorder	F40–F48 Neurotic, stress-related & somatoform	0.70	0.49
Anxiety OC&SR	PTSD	F43.1 Posttraumatic stress disorder	F40–F48 Neurotic, stress-related & somatoform	0.69	0.41
Anxiety OC&SR	SocPhob	F40.1 Social phobia	F40–F48 Neurotic, stress-related & somatoform	0.73	0.48
Dissociative & somatoform	Convers	F44 Conversion disorders	F40–F48 Neurotic, stress-related & somatoform	0.62	0.49
Dissociative & somatoform	Dissoc	F44 Dissociative disorders	F40–F48 Neurotic, stress-related & somatoform	0.62	0.41
Dissociative & somatoform	Hypochon	F45.2 Hypochondriacal disorder	F40–F48 Neurotic, stress-related & somatoform	0.59	0.46
Dissociative & somatoform	PersPain	F45.4 Persistent somatoform pain disorder	F40–F48 Neurotic, stress-related & somatoform	0.59	0.49
Dissociative & somatoform	Somatiz	F45.0 Somatization disorder	F40–F48 Neurotic, stress-related & somatoform	0.59	0.51
Mood	BDI	F31 Bipolar I disorder	F30–F39 Mood [affective]	0.70	0.63
Mood	BDII	F31.8 Bipolar II disorder	F30–F39 Mood [affective]	0.70	0.63

(Continues)

TABLE 1 (Continued)

Cluster[a]	Abbreviation	Diagnostic code and category[b]	Chapter[b]	Clinical utility[c]	Cohesiveness[a]
Mood	MDD	F33 Depressive disorder (major)	F30–F39 Mood [affective]	0.72	0.61
Neurocognitive & "organic"	AlzDem	F00 Alzheimer's dementia	F00–F09 Organic, including symptomatic	0.67	0.59
Neurocognitive & "organic"	Delir	F05 Delirium	F00–F09 Organic, including symptomatic	0.69	0.55
Neurocognitive & "organic"	VasDem	F01 Vascular dementia	F00–F09 Organic, including symptomatic	0.64	0.59
Neurodevelopmental	ADHD	F90 Attention deficit-hyperactivity (hyperkinetic) disorder	F90–F98 Onset usually occurring in child & adolesc	0.60	0.48
Neurodevelopmental	Asp	F84.5 Asperger's syndrome	F80-F89 Disorders of psychological development	0.57	0.54
Neurodevelopmental	Autis	F84.0 Autistic disorder	F80–F89 Disorders of psychological development	0.56	0.55
Neurodevelopmental	IntellDis	F7 Intellectual disability (mental retardation)	F70–F79 Mental retardation	0.71	0.45
Other bodily	Anorex	F50.0 Anorexia nervosa	F50–F59 Associated with physiol and physical factors	0.66	0.19
Other childhood	Conduct	F91 Conduct disorder	F90–F98 Onset usually occurring in child & adolesc	0.58	0.32
Other childhood	ODD	F91.3 Oppositional defiant disorder	F90–F98 Onset usually occurring in child & adolesc	0.58	0.35
Other childhood	Tic	F95 Tic disorders	F90–F98 Onset usually occurring in child & adolesc	0.70	0.20
Personality	AntiPD	F60.2 Antisocial (dissocial) personality disorder	F60–F69 Adult personality & behavior	0.64	0.60
Personality	BorderPD	F60.31 Borderline personality disorder	F60–F69 Adult personality and behavior	0.56	0.61
Schizophrenia S&OPP	BrfPsychot	F23 Acute and transient (brief) psychotic disorder	F20–F29 Schizophrenia, schizotypal and delusional	0.65	0.55
Schizophrenia S&OPP	Delus	F22 Delusional disorder	F20–F29 Schizophrenia, schizotypal & delusional	0.65	0.57

TABLE 1 (Continued)

Cluster[a]	Abbreviation	Diagnostic code and category[b]	Chapter[b]	Clinical utility[c]	Cohesiveness[a]
Schizophrenia S&OPP	Schizoaff	F25 Schizoaffective disorder	F20–F29 Schizophrenia, schizotypal & delusional	0.58	0.52
Schizophrenia S&OPP	Schizophr	F20 Schizophrenia	F20–F29 Schizophrenia, schizotypal & delusional	0.68	0.59
Schizophrenia S&OPP	Schizotyp	F21 Schizotypal disorder	F20–F29 Schizophrenia, schizotypal & delusional	0.57	0.45
Sexual	SexDysf	F52 Sexual dysfunction	F50–F59 Associated with physiol & physical factors	0.61	0.44
Substance	AlcDep	F10.2 Alcohol dependence	F10–F19 Due to psychoactive substance use	0.70	0.68
Substance	CanAbu	F12.1 Cannabinoid abuse	F10–F19 Due to psychoactive substance use	0.68	0.67
Substance	CocDep	F14.2 Cocaine dependence	F10–F19 Due to psychoactive substance use	0.69	0.68
Substance	OpiDep	F11.2 Opioid dependence	F10–F19 Due to psychoactive substance use	0.66	0.68
Substance	SolvAbu	F18.1 Abuse of volatile solvents (inhalants)	F10–F19 Due to psychoactive substance use	0.67	0.67

Note: Data are presented in a transformed metric from 0 (low utility/cohesiveness) to 1 (high utility/cohesiveness).

Abbreviations: child & adolesc, childhood & adolescence; OC&SR, obsessive-compulsive, & stress-related; physiol, physiological; S&OPP, spectrum & other primary psychotic.

[a]From the clinicians' taxonomy study (Reed et al., 2013).

[b]From ICD-10 (World Health Organization, 1992b).

[c]From the global psychiatrists survey study (Reed et al., 2011).

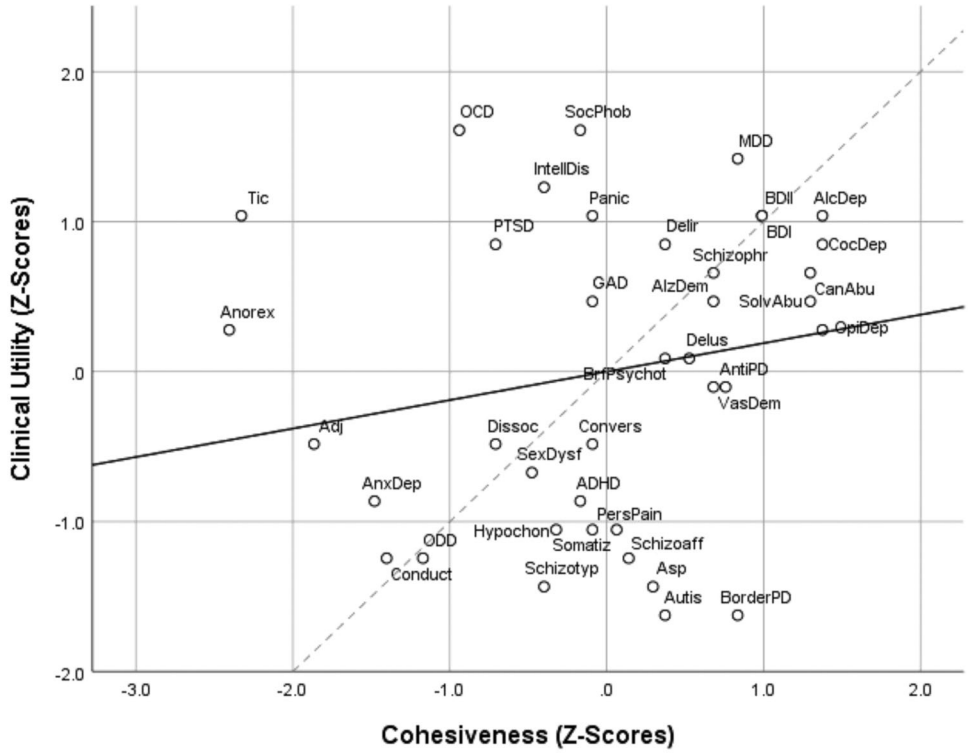

FIGURE 1 Scatterplot for the correlation between clinical utility and taxonomic cohesiveness in mental disorder categories. See Table 1 for abbreviations. The dark solid line represents the observed correlation ($r = 0.19$). The gray dashed line serves as a reference line for hypothetical positive correlation ($r = 1$)

opposite extreme (top left), disorders such as OCD, Tic Disorders, and Anorexia Nervosa show relatively poor cohesiveness but above-average clinical utility. The top right quadrant includes several categories with good fit and good utility, predominantly characterized by the substance use and mood disorders. Finally, the bottom left quadrant shows disorders that trended toward low utility and low cohesiveness, including Oppositional Defiant Disorder, Conduct Disorder, and Mixed Anxiety and Depressive Disorder.

4 | DISCUSSION

This study offers the first empirical examination of the association between taxonomy and utility in mental disorder diagnostic classification. Following the logic of the ICD-11 development, we expected to see a positive association between the taxonomic fit and the clinical utility of mental disorders. Results were generally consistent with this hypothesis. There was a positive overall correlation between taxonomic fit and clinical utility, although the coefficient estimate was somewhat smaller in magnitude than expected and not statistically different from zero based on its 95% CIs. Overall, diagnostic categories that fit well within clinicians' conceptual landscape of psychopathology tended to be the same categories that had higher clinical utility (i.e., rated by clinicians as being easy to use and accurately describing patients). Closer examination yielded further interesting results, discussed below.

It seems likely that the *true* taxonomy-utility correlation may be larger than the observed coefficient of $r = 0.19$ [−0.06, 0.43]. Follow-up analyses suggested that the effect could be closer to $r = 0.34$ [0.05, 0.58] or even $r = 0.58$ [0.34, 0.79], but that the strength of the association was attenuated by a few outliers in our data. Notably, these

outliers were identified through objective detection methods at varying thresholds; therefore, their removal would not necessarily be expected to amplify the correlation as it did. In fact, an interesting dose–response pattern was observed: with every step toward more conservative data inclusion criteria (first, all data; second, all data within $M \pm 2\ SD$; third, all data within $M \pm 1.5\ SD$), the Pearson's r rose to the next threshold for effect size interpretation (from small to medium to large, respectively). It should be reiterated that these analyses were exploratory, so the higher correlations should only be interpreted alongside the full-data correlation. At the same time, another caveat is warranted: our data-analytic "sample" was 39 categories drawn from the much larger "population" of diagnostic constructs that exist in mental health. Further research drawing new samples and different types of data from this larger population would shed further light on the phenomenon examined here; our outlier analyses suggest a somewhat stronger positive correlation might be found.

The positive association between taxonomy and utility aligns with the rationale underlying the ICD-11 research program. About a decade ago, it was argued that clinical utility should be prioritized in the revision of ICD-11 by bringing the structure of the formal taxonomy into closer alignment with the structure of clinicians' cognitive taxonomies (International Advisory Group for the Revision of ICD-10 Mental and Behavioral Disorders, 2011; Reed, 2010). The present results suggest that, indeed, the diagnostic categories that actual clinicians perceived as having poorer clinical utility tend to be those that do not fit as well with how clinicians think about them in relation to the broader psychiatric nosology. This indicates that clinical researchers should attend to clinicians' insights into the phenomenology of mental and behavioral disorders they encounter. Here it is important to acknowledge the distinction between correlation and causation. That is, supposing that the developers of ICD-11 were successful in making diagnostic categories with low cohesion indices in ICD-10 somehow "fit" better into the structure of the ICD-11, this would not necessarily mean that these same diagnostic categories would show corresponding improvements in clinical utility (a cause–effect relationship). Indeed, it could be that some categories that do not fit well with other diagnoses are useful because they are dissimilar from other conditions and thereby fill a gap in symptom coverage.

4.1 | Findings concerning specific diagnostic categories

Examination of individual data points was informative in its own right. Categories with *low utility but above-average cohesiveness* included Asperger's Syndrome, Autistic Disorder, and Borderline Personality Disorder. It may be that these disorders seem relatively straightforward to classify—ostensibly falling among the "child" or "personality" disorders—but that they pose challenges clinically, especially in terms of their heterogeneity and boundaries with normality. Recent diagnostic revisions and debates are consistent with these findings. In ICD-11 and DSM-5, Asperger Syndrome was subsumed into Autism Spectrum Disorder (ASD), with the rationale that ASD is a more heterogeneous diagnosis that falls on a spectrum of symptomatic clinical presentation (Clark et al., 2017; Lord & Jones, 2012). Similarly, it has long been debated whether Borderline Personality Disorder is more beneficially conceptualized from a categorical or dimensional perspective (Clark, 1993; Griffiths, 2011; Verheul, 2005). The ICD-11 includes a unique mechanism for denoting a "borderline pattern" among personality disorder symptoms despite moving from a categorical to a dimensional nosology (Reed, 2018; Tyrer et al., 2019). This qualifier was included to enhance clinical utility and identify individuals who may respond to certain psychotherapeutic treatments, based largely on demand from the clinical community.

Conversely, categories with *low cohesiveness but above-average utility*—including Tic Disorders, Anorexia Nervosa, and OCD—may be accurate and useful but hard to classify. These conditions are relatively distinctive, with hallmark patterns of both emotional distress and repeated behaviors, following a somewhat chronic course, and sometimes with a particular treatment indicated (American Psychiatric Association, 2013; World Health Organization, 2020). Such characteristics may offer clinical utility. But in each case, research has yielded competing views and shifts regarding where they should be placed. Notably, OCD was moved from Anxiety Disorders to a

separate section for Obsessive-Compulsive and Related Disorders in both ICD-11 and DSM-5, reflecting changes in scientific understanding (Marras et al., 2016; Stein et al., 2016). But the clinical knowledge and usage of this entity seems more stable. For OCD, first-line effective treatments continue to include exposure-based cognitive-behavioral therapies, very similar to anxiety disorders, even if not classified among them (Franklin & Foa, 2011; McKay et al., 2015). Similarly, Tic Disorder's high utility might reflect a clearly identifiable clinical presentation and indicated behavioral treatment (Essoe et al., 2019), in spite of, not because of, its taxonomic fit. The poor fit might be related to its heterogeneity and comorbidities including OCD, ASD, ADHD, and Learning Disabilities (Murphy et al., 2013). Similar points could be raised for Anorexia Nervosa, but its placement as an outlier may simply be an artifact of not including other eating disorders in the original study (Reed et al., 2013).

Most of the diagnostic categories in the scatterplot were not exceptions or outliers, and their placement on the two-dimensional figure seem to offer useful insights about the relation between clinical utility and taxonomic fit. The *high cohesiveness–high utility* categories included essentially all present mood disorder categories (both depressive and bipolar), substance use disorders, Schizophrenia, and Alzheimer's Dementia. These constructs could be interpreted (with a grain of salt) as ones that are "working well" for clinicians in terms of accuracy, ease of use, and taxonomic placement. They also happen to be quite salient, including unique features that are not present in other disorders (e.g., the cardinal symptoms of manic episodes, the positive symptoms of psychosis) as well as clear boundaries with normality (e.g., dependencies in substance use disorders, episodic patterns in mood disorders, degenerative patterns in Alzheimer's). Interestingly, and perhaps relatedly, these are some of the oldest syndromes in the history of mental illness. Dementia praecox (the predecessor to schizophrenia), manic-depression (a common ancestor of modern depressive and bipolar disorders), and Alzheimer's were all described by Emil Kraepelin and colleagues over a century ago (Engstrom & Kendler, 2015).

Lastly, the *low cohesiveness–ow utility* categories include Oppositional Defiant Disorder, Conduct Disorder, Mixed Anxiety and Depressive Disorder, and Adjustment Disorders. These conditions may not fit "cleanly" into a single disorder block and may also be unfamiliar or challenging for some clinicians to use. For example, Oppositional Defiant Disorder and Conduct Disorder are presumably viewed by most clinicians as child and adolescent problems that involve specific patterns of externalizing behavior; in reality, however, these two disorders can be present across the lifespan and encompass a large, heterogeneous array of very distinct problems (e.g., arguing, stealing, lying, assault, cruelty to animals, irritability/anger; Evans et al., 2017; Fairchild et al., 2019). Further, the high prevalence of disruptive behavior disorders may lead many clinicians to endorse some level of familiarity with them, but such "childhood conditions" are complex and challenging, overlapping with normative development, warranting specialized care from youth mental health professionals internationally when available (Roberts et al., 2016). Similarly, anxious/depressive and adjustment disorders encompass an array of different presentations (e.g., anxiety, depressed mood, stress) relating to affective distress in some form. Adjustment Disorder has been treated as a "catchall" category assigned to individuals who did not reach thresholds for disorders but needed treatment and reimbursement thereof (Bachem & Casey, 2018, p. 244). Likewise, Mixed Anxiety and Depressive Disorder has attracted some criticism but seems useful in clinical practice, especially primary care, for promoting access to care and early treatment of persons with anxious and depressive symptoms (Möller et al., 2016).

4.2 | Limitations, strengths, and implications

This study has some limitations, all largely related to it being a secondary data analysis. This affected what disorders were included and excluded. It also led us to focus only on two specific components of clinical utility—ease of use and goodness of fit—which clinicians are especially well-situated to report because they see patients with mental disorders on a routine basis. There are other components of clinical utility that this study did not address (Reed, 2010). Future research is needed to examine value for treatment, communication, and policy/societal uses. Of course, multiple diagnostic classification systems exist, and our focus on ICD is consistent with the

practices of the great majority of mental health professionals according to global clinician surveys (Evans et al., 2013; First et al., 2018; Reed et al., 2011). Examination of differences between alternative systems (e.g., ICD-11 vs. DSM-5) was beyond the scope of this analysis and has been addressed regarding specific questions in field trials (e.g., Evans et al., 2020). Although not a limitation, it is worth noting that our participants were clinicians, not patients. Research with patients is of course necessary for advancing the understanding and treatment of mental disorders. But our findings show the necessity and importance of research with clinicians, and the need for further such work focusing on clinicians who provide care using ICD-11 and DSM-5.

Strengths of the present study are also related to strengths of the original investigations. Both studies were truly global in scope, yielding large samples of mental health professionals representing different languages, cultures, and countries. Indeed, the scale and substance of this study is without precedent and could not have occurred outside of WHO's ICD-11 development. Despite our limitation of a *small sample* of only 39 mental disorder categories—which restricted our power to detect smaller effect sizes as statistically significant—the data gathered on these 39 categories were in fact high-quality measurements drawn from *very large global samples*. As such, these variables functioned as appropriate indicators of how diagnostic entities are conceptualized in regular clinical practice. Moreover, these results have broad generalizability, reflecting the inherently global nature of science and health.

Practical implications are also apparent. Diagnostic categories' placements on the utility × fit plane could be treated as draft research and training agenda, including targeted education efforts for clinicians at all levels of training and experience (Roberts et al., 2016). From this premise, it would follow that training might be especially helpful for categories in the bottom and left quadrants of Figure 1, though with different emphases. Categories with low utility (e.g., ASD, Borderline Personality Disorder, related concerns) might warrant training focused on the presentation, assessment, and treatment of these conditions, to help clinicians use them with greater accuracy and ease. Disorders with low cohesiveness (e.g., Tic Disorders, Anorexia Nervosa, related concerns) might require training on their nosological placement, relations to co-occurring problems, and differential diagnosis. And for conditions low in both utility and cohesiveness (e.g., Oppositional Defiant Disorder, Conduct Disorder, related concerns), all these types of training could be helpful. Alternatively, patients with these low-utility or low-fit conditions might be best served by a referral for a more specialized provider or a more intensive level of care, when such options are available. There are similar implications for practicing clinicians. For instance, the disorders noted in this paragraph might be flagged as referral problems that warrant a more thorough clinical assessment, a consultation with a colleague, or a referral to a secondary or tertiary provider. At the same time, findings in studies like this one are intractably affected by the pros and cons of categorical classification more generally (Clark et al., 2017). For example, diagnostic heterogeneity, comorbidity, and boundaries with normality were not measured in this study; yet, when considering individual data points, such characteristics of diagnoses seemed relevant to the correspondence between taxonomic fit and clinical utility. These issues warrant more explicit consideration in future research.

Decades have passed since the prior ICD revision, and it will likely be several more years to the next rounds of diagnostic revisions. Periodic revisions are necessary to help advance mental health research and practice in line with the best available evidence. It is therefore important that research on diagnostic classification continues to be carried out, even in the intervening time between major revision efforts. The present study suggests that there is a positive association between the taxonomic fit of mental disorders and their clinical utility among global clinicians. These findings align with the notion that improving the taxonomic fit of specific diagnostic constructs might be one tractable strategy for improving clinical utility in general, with some specific diagnostic categories being identified as targeted areas for education, practice, research, and classification. In these ways, the present study makes a unique and important contribution to the evidence base on the diagnostic classification of mental disorders. Additional research is needed to refine the measurement of clinical utility and to advance the understanding and applications of clinicians' taxonomies of mental disorders.

ACKNOWLEDGMENTS

This article draws data from two previously published studies: Reed et al. (2011, 2013). The first author gratefully acknowledges support from AIM for Mental Health (Clinical Science Fellowship) and National Institute of Mental Health (Clinical Loan Repayment Program) during preparation of this manuscript. Michael C. Roberts was a member of, and Spencer C. Evans and Jared W. Keeley were consultants to, the Field Studies Coordination Group for ICD-11 Mental and Behavioral Disorders. Geoffrey M. Reed is a member of the WHO Secretariat. Unless specifically stated, the views expressed in this article are those of the authors and do not represent the official policies or positions of WHO.

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1002/jclp.23125.

DATA AVAILABILITY STATEMENT

Data sharing not applicable—no new data generated.

ORCID

Spencer C. Evans http://orcid.org/0000-0001-8644-817X

REFERENCES

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)*. American Psychiatric Publications.

Bachem, R., & Casey, P. (2018). Adjustment disorder: A diagnosis whose time has come. *Journal of Affective Disorders*, *227*, 243–253.

Clark, L. A. (1993). Personality disorder diagnosis: Limitations of the Five-Factor Model. *Psychological Inquiry*, *4*, 100–104.

Clark, L. A., Cuthbert, B., Lewis-Fernández, R., Narrow, W. E., & Reed, G. M. (2017). Three approaches to understanding and classifying mental disorder: ICD-11, DSM-5, and the National Institute of Mental Health's Research Domain Criteria (RDoC). *Psychological Science in the Public Interest*, *18*(2), 72–145.

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*(1), 7–29.

Engstrom, E. J., & Kendler, K. S. (2015). Emil Kraepelin: Icon and reality. *American Journal of Psychiatry*, *172*(12), 1190–1196.

Essoe, J. K. Y., Grados, M. A., Singer, H. S., Myers, N. S., & McGuire, J. F. (2019). Evidence-based treatment of Tourette's disorder and chronic tic disorders. *Expert Review of Neurotherapeutics*, *19*(11), 1103–1115.

Evans, S. C., Burke, J. D., Roberts, M. C., Fite, P. J., Lochman, J. E., de la Peña, F. R., & Reed, G. M. (2017). Irritability in child and adolescent psychopathology: An integrative review for ICD-11. *Clinical Psychology Review*, *53*, 29–45. https://doi.org/10.1016/j.cpr.2017.01.004

Evans, S. C., Reed, G. M., Roberts, M. C., Esparza, P., Watts, A. D., Correia, J. M., & Saxena, S. (2013). Psychologists' perspectives on the diagnostic classification of mental disorders: Results from the WHO-IUPsyS global survey. *International Journal of Psychology*, *48*(3), 177–193.

Evans, S. C., Roberts, M. C., Keeley, J. W., Rebello, T. J., Peña, F., Lochman, J. E., Burke, J. D., Fite, P. J., Ezpeleta, L., Matthys, W., Youngstrom, E. A., Matsumoto, C., Andrews, H. F., Elena Medina-Mora, M., Ayuso-Mateos, J. L., Khoury, B., Kulygina, M., Robles, R., Sharan, P., … Reed, G. M. (2020). Diagnostic classification of irritability and oppositionality in youth: A global field study comparing ICD-11 with ICD-10 and DSM-5. *Journal of Child Psychology and Psychiatry*. (Advance online publication). https://doi.org/10.1111/jcpp.13244

Fairchild, G., Hawes, D. J., Frick, P. J., Copeland, W. E., Odgers, C. L., Franke, B., Freitag, C. M., & De Brito, S. A. (2019). Conduct disorder. *Nature Reviews Disease Primers*, *5*(1), 1–25.

First, M. B., Pincus, H. A., Levine, J. B., Williams, J. B. W., Ustun, B., & Peele, R. (2004). Clinical utility as a criterion for revising psychiatric diagnoses. *American Journal of Psychiatry*, *161*(6), 946–954.

First, M. B., Rebello, T. J., Keeley, J. W., Bhargava, R., Dai, Y., Kulygina, M., Matsumoto, C., Robles, R., Stona, A. C., & Reed, G. M. (2018). Do mental health professionals use diagnostic classifications the way we think they do? A global survey. *World Psychiatry*, *17*(2), 187–195.

First, M. B., Reed, G. M., Hyman, S. E., & Saxena, S. (2015). The development of the ICD-11 clinical descriptions and diagnostic guidelines for mental and behavioural disorders. *World Psychiatry*, *14*, 82–90.

First, M. B., Wolfgang, G., Maj, M., Stein, D., Kogan, C. S., Saunders, J. B., & Reed, G. M.An organization- and category-level comparison of diagnostic requirements for mental disorders in ICD-11 and DSM-5. *World Psychiatry*. (In Press).

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

Flanagan, E. H., & Blashfield, R. K. (2007a). Clinicians' folk taxonomies of mental disorders. *Philosophy, Psychiatry, & Psychology*, *14*(3), 249–269.

Flanagan, E. H., & Blashfield, R. K. (2007b). Should clinicians' views of mental illness influence the DSM? *Philosophy, Psychiatry, & Psychology*, *14*(3), 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*(6), 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*(3), 693–698.

Flanagan, E., Keeley, J., & Blashfield, R. (2012). Do clinicians conceptualize DSM-IV disorders hierarchically? *Journal of Clinical Psychology*, *68*(6), 620–630.

Franklin, M. E., & Foa, E. B. (2011). Treatment of obsessive compulsive disorder. *Annual Review of Clinical Psychology*, *7*, 229–243.

Griffiths, M. (2011). Validity, utility and acceptability of borderline personality disorder diagnosis in childhood and adolescence: Survey of psychiatrists. *The Psychiatrist*, *35*, 19–22.

Halsey, L. G., Curran-Everett, D., Vowler, S. L., & Drummond, G. B. (2015). The fickle P value generates irreproducible results. *Nature Methods*, *12*(3), 179–185.

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., Sanislow, C., & Wang, P. (2010). Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. *American Journal of Psychiatry*, *167*(7), 748–751.

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*(2), 86–92.

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 Behavioural Disorders. *American Psychologist*, *71*, 3–16.

Kendall, R., & Jablensky, A. (2003). Distinguishing between the validity and utility of psychiatric diagnoses. *American Journal of Psychiatry*, *160*(1), 4–12.

Kupfer, D. J., First, M. B., & Regier, D. A. (Eds.). (2002). *A research agenda for DSM-V*. American Psychiatric Association.

Kupfer, D. J., Regier, D. A., & Kuhl, E. A. (2008). On the road to DSM-V and ICD-11. *European Archives of Psychiatry and Clinical Neuroscience*, *258*(Suppl 5), 2–6.

Lord, C., & Jones, R. M. (2012). Annual research review: Re-thinking the classification of autism spectrum disorders. *Journal of Child Psychology and Psychiatry*, *53*, 490–509.

Marras, A., Fineberg, N., & Pallanti, S. (2016). Obsessive compulsive and related disorders: Comparing DSM-5 and ICD-11. *CNS Spectrums*, *21*(4), 324–333.

McKay, D., Sookman, D., Neziroglu, F., Wilhelm, S., Stein, D. J., Kyrios, M., Matthews, K., & Veale, D. (2015). Efficacy of cognitive-behavioral therapy for obsessive–compulsive disorder. *Psychiatry Research*, *225*(3), 236–246.

Mellsop, G., Dutu, G., Pereira, M., Nagamine, M., Shinfuku, N., & Banzato, C. (2007). An international study of the views of psychiatrists on present and preferred characteristics of classifications of psychiatric disorders. *International Journal of Mental Health*, *36*(4), 17–25.

Möller, H. J., Bandelow, B., Volz, H. P., Barnikol, U. B., Seifritz, E., & Kasper, S. (2016). The relevance of 'mixed anxiety and depression' as a diagnostic category in clinical practice. *European Archives of Psychiatry and Clinical Neuroscience*, *266*(8), 725–736.

Mullins-Sweatt, S. N., & Widiger, T. A. (2009). Clinical utility and *DSM-V*. *Psychological Assessment*, *21*(3), 302–312.

Murphy, T. K., Lewin, A. B., Storch, E. A., & Stock, S. (2013). Practice parameter for the assessment and treatment of children and adolescents with tic disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, *52*(12), 1341–1359.

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*(6), 457–464.

Reed, G. M. (2018). Progress in developing a classification of personality disorders for ICD-11. *World Psychiatry*, *17*, 227–229.

Reed, G. M., Correia, J. M., Esparza, P., Saxena, S., & Maj, M. (2011). The WPA-WHO global survey of psychiatrists' attitudes toward mental disorders classification. *World Psychiatry*, *10*, 118–131.

Reed, G. M., Keeley, J. W., Rebello, T. J., First, M. B., Gureje, O., Ayuso-Mateos, J. L., Kanba, S., Khoury, B., Kogan, C. S., Krasnov, V. N., Maj, M., de Jesus Mari, J., Sharan, P., Stein, D. J., Zhao, M., Akiyama, T., Andrews, H. F., Asevedo, E., … Medina-Mora, M. E. (2018). Clinical utility of ICD-11 diagnostic guidelines for high-burden mental disorders: Results from mental health settings in 13 countries. *World Psychiatry*, *17*(3), 306–315.

Reed, G. M., Roberts, M. C., Keeley, J., Hooppell, C., Matsumoto, C., Sharan, P., Robles, R., Carvalho, H., Wu, C., Gureje, O., Leal-Leturia, I., Flanagan, E. H., Correia, J. M., Maruta, T., Ayuso-Mateos, J. L., de Jesus Mari, J., Xiao, Z., Evans, S. C., … Medina-Mora, M. E. (2013). Mental health professionals' natural taxonomies of mental disorders: Implications for the clinical utility of the ICD-11 and the DSM-5. *Journal of Clinical Psychology*, *69*, 1191–1212.

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

Regier, D., Narrow, W., Kuhl, E., & Kupfer, D. (2009). The conceptual development of DSM-V. *American Journal of Psychiatry*, *166*, 645–650.

Roberts, M. C., Kanine, R. M., Amaro, C. M., Evans, S. C., Blossom, J. B., & Garcia, A. M. (2016). International education and training for clinical child and adolescent psychology. *South African Journal of Psychology*, *46*(1), 9–24.

Roberts, M. C., Reed, G. M., Medina-Mora, M. E., Keeley, J. W., Sharan, P., Johnson, D. K., Mari, J. J., Ayuso-Mateos, J. L., Gureje, O., Xiao, Z., Maruta, T., Khoury, B., Robles, R., & Saxena, S. (2012). A global clinicians' map of mental disorders to improve ICD-11: Analysing meta-structure to enhance clinical utility. *International Review of Psychiatry*, *24*(6), 578–590.

Sartorius, N. (1993). Progress toward achieving a common language in psychiatry: Results from the field trial of the clinical guidelines accompanying the WHO classification of mental and behavioral disorders in ICD-10. *Archives of General Psychiatry*, *50*, 115–124.

Shaffer, D. (1996). A participant's observations: Preparing DSM-IV. *The Canadian Journal of Psychiatry/La Revue Canadienne De Psychiatrie*, *41*(6), 325–329.

Stein, D. J., Kogan, C. S., Atmaca, M., Fineberg, N. A., Fontenelle, L. F., Grant, J. E., Matsunaga, H., Reddy, Y. C. J., Simpson, H. B., Thomsen, P. H., van den Heuvel, O. A., Veale, D., Woods, D. W., & Reed, G. M. (2016). The classification of obsessive–compulsive and related disorders in the ICD-11. *Journal of Affective Disorders*, *190*, 663–674.

Suzuki, Y., Takahashi, T., Nagamine, M., Zou, Y., Cui, J., Han, B., Park, J. I., Hwu, H. G., Chen, C. C., Lin, C. C., & Shinfuku, N. (2010). Comparison of psychiatrists' views on classification of mental disorders in four east Asian countries/area. *Asian Journal of Psychiatry*, *3*(1), 20–25.

Tyrer, P., Mulder, R., Kim, Y. R., & Crawford, M. J. (2019). The development of the ICD-11 classification of personality disorders: An amalgam of science, pragmatism, and politics. *Annual Review of Clinical Psychology*, *15*, 481–502. https://doi.org/10.1146/annurev-clinpsy-050718-095736

Verheul, R. (2005). Clinical utility of dimensional models for personality pathology. *Journal of Personality Disorders*, *19*, 283–302.

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

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

World Health Organization. (2019). World health assembly update 25 may 2019: International statistical classification of diseases and related health problems (ICD-11). https://www.who.int/news-room/detail/25-05-2019-world-health-assembly-update

World Health Organization. (2020). *Classifications: International classification of diseases, 11th revision (ICD-11)*. http://www.who.int/classifications/icd/en/

Zielasek, J., Freyberger, H. J., Janner, M., Kapfhammer, H. P., Sartorius, N., Stieglitz, R. D., & Gaebel, W. (2010). Assessing the opinions and experiences of German-speaking psychiatrists regarding necessary changes for the 11th revision of the mental disorders chapter of the International Classification of Disorders (ICD-11). *European Psychiatry*, *25*(8), 437–442.

How to cite this article: Evans, S. C., Roberts, M. C., Guler, J., Keeley, J. W., Reed, G. M. (2021). Taxonomy and utility in the diagnostic classification of mental disorders. *Journal of Clinical Psychology*. 77, 1921–1936. https://doi.org/10.1002/jclp.23125