Postprint of: Evans, S. C., Roberts, M.C. Keeley, J.W., Rebello, T.J., de la Peña, F., Lochman, J.E., Burke, J.D., Fite, P.J., Ezpeleta, L., Matthys, W., Youngstrom, E.A., Matsumoto, C., Andrews, H.F., Medina-Mora, M.E., Ayuso-Mateos, J.L., Khoury, B., Kulygina, M., Robles, R., Sharan, P., Zhao, M., & Reed, G.M. (2021). 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, 62, 303-312 doi: 10.1111/jcpp.13244.

Diagnostic Classification of Irritability and Oppositionality in Youth: A Global Field Study Comparing ICD-11 with ICD-10 and DSM-5

Spencer C. Evans[1,*], Michael C. Roberts[2], Jared W. Keeley[3], Tahilia J. Rebello[4,5], Francisco de la Peña[6], John E. Lochman[7], Jeffrey D. Burke[8], Paula J. Fite[2], Lourdes Ezpeleta[9], Walter Matthys[10], Eric A. Youngstrom[11], Chihiro Matsumoto[12], Howard F. Andrews[13], María Elena Medina-Mora[6], José L. Ayuso-Mateos[14], Brigitte Khoury[15], Mayya Kulygina[16], Rebeca Robles[6], Pratap Sharan[17], Min Zhao[18], & Geoffrey M. Reed[4,19]

[1]Harvard University, Cambridge, MA, USA

[2]University of Kansas, Lawrence, KS, USA

[3]Virginia Commonwealth University, Richmond, VA, USA

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

[5]Research Foundation for Mental Hygiene, New York, NY, USA

[6]Instituto Nacional de Psiquiatría Ramón de la Fuente Muñiz, Mexico City, Mexico

[7]University of Alabama, Tuscaloosa, AL, USA

[8]University of Connecticut, Storrs, CT, USA

[9]Universitat Autonoma de Barcelona, Barcelona, Spain

[10]Utrecht University, Utrecht, The Netherlands

[11]University of North Carolina, Chapel Hill, NC, USA

[12]Japanese Society of Psychiatry and Neurology, Tokyo, Japan

[13]Columbia University Medical Center, New York, NY, USA

[14]Universidad Autónoma de Madrid, Madrid, Spain

[15]American University of Beirut Medical Center, Beirut, Lebanon

[16]Alekseev Mental Health Clinic №1, Moscow, Russian Federation

[17]All India Institute of Medical Sciences, New Delhi, India

[18]Shanghai Jiao Tong University School of Medicine, Shanghai, People's Republic of China

[19]World Health Organization, Geneva, Switzerland

*Correspondence: scevans@fas.harvard.edu. Spencer Evans is now affiliated with the Department of Psychology at the University of Miami.

ABSTRACT

Background: Severe irritability has become an important topic in child and adolescent mental health. Based on the available evidence and on public health considerations, WHO classified chronic irritability within Oppositional Defiant Disorder (ODD) in ICD-11, a solution markedly different from DSM-5's (i.e., the new childhood mood diagnosis, Disruptive Mood Dysregulation Disorder [DMDD]) and from ICD-10's (i.e., ODD as one of several conduct disorders without attention to irritability). In this study, we tested the accuracy with which a global, multilingual, multidisciplinary sample of clinicians were able to use the ICD-11 classification of chronic irritability as compared to the ICD-10 and DSM-5 approaches. **Methods:** Clinicians (*N*=196) from 48 countries participated in an internet-based field study in English, Spanish, or Japanese, and were randomized to review and use one of the three diagnostic systems. Through experimental manipulation of validated clinical vignettes, we evaluated how well clinicians in each condition could identify chronic irritability vs. non-irritable oppositionality, episodic bipolar disorder, dysthymic depression, and normative irritability. **Results:** Compared to ICD-10 and DSM-5, ICD-11 led to more accurate identification of severe irritability and better differentiation from boundary presentations. Participants using DSM-5 largely failed to apply the DMDD diagnosis when it was appropriate, and they more often applied psychopathological diagnoses to developmentally normative irritability. **Conclusions:** The formulation of irritability and oppositionality put forth in ICD-11 shows evidence of clinical utility, supporting accurate diagnosis. Global mental health clinicians can readily identify ODD both with and without chronic irritability.

Keywords: International Classification of Diseases (ICD-11), Oppositional defiant disorder (ODD); mood dysregulation; irritability; child and adolescent mental health

INTRODUCTION

Irritability in children and adolescents (herein "youth") has emerged as a significant clinical and public health concern for which assessment and treatment options are limited. Defined as an elevated proneness to anger, irritability has manifestations ranging from normative emotions to chronically irritable mood and aggressive outbursts (Stringaris et al., 2018). Irritability occurs in over a dozen mental disorders including Oppositional Defiant Disorder (ODD), where it is chronic and often a core part of the presentation (Evans et al., 2017). Longitudinal studies show that irritability does not specifically predict bipolar disorder (BD) or conduct disorder (CD), but it does predict ODD, depression, anxiety disorders, suicidality, and poor functional outcomes (Brotman, Kircanski, & Leibenluft, 2017; Evans et al., 2017; Vidal-Ribas et al., 2016). Thus, it is important that diagnostic classification systems accurately identify and characterize severely irritable youth in need of clinical attention. Chronic irritability was therefore given close attention in the development of the Mental, Behavioural, and Neurodevelopmental Disorders (MBND) chapter in the Eleventh Revision of the World Health Organization's (WHO's) *International Classification of Diseases and Related Health Problems* (ICD-11). The purpose of the present study was to evaluate the utility of the ICD-11 formulation of irritability and oppositionality for global mental health applications.

In May 2019, the World Health Assembly approved ICD-11 as the global standard for collection and reporting of health information by WHO's 194 member nations (Reed et al., 2019). The development of ICD-11 MBND was a massive and globally influential undertaking coordinated by the WHO Department of Mental Health and Substance Abuse (First et al., 2015; International Advisory Group, 2011; Reed, 2010; Reed et al., 2019). As a part of this effort, the Department developed detailed diagnostic guidelines for clinician use and implemented a program of global field studies to assess the reliability and clinical utility of these guidelines (First et al., 2015; Keeley et al., 2016). At the time of this study, the ICD-11 draft guidelines had not yet been finalized, allowing a window for potential refinement based on field study results.

Early in ICD-11 development, clinical utility and global applicability were identified as guiding principles based on the idea that improvements in these domains would foster improvements in clinical

service delivery and resulting health data (International Advisory Group, 2011; Reed, 2010). To this end, an international group of experts in disruptive mood and behavior problems (e.g., aggression, oppositionality, mood disorders, irritability) reviewed the literature and proposed changes intended to (a) improve communication among users (e.g., practitioners, patients, families, administrators); (b) foster conceptualization and understanding of Disruptive Behaviour and Dissocial Disorders (DBDDs); (c) accurately and easily describe actual clinical presentations; (d) assist with clinical management; (e) enhance clinical outcomes (Reed, 2010); and (f) enhance applicability of these changes by multidisciplinary clinicians in diverse settings across the globe.

It was the view of this and other ICD-11 expert working groups that existing evidence argued against introducing chronic irritability as a new standalone disorder. Instead, it was proposed that irritability be identified as a feature of ODD (Evans et al., 2017; Lochman et al., 2015). This decision was justified in part by evidence showing that irritability is a major subdimension of ODD symptoms (e.g., Burke et al., 2014; Ezpeleta et al., 2012; Rowe et al., 2010; Stringaris & Goodman, 2009). Converging research on ODD-irritability and alternative models of chronic irritability (e.g., severe mood dysregulation [SMD]; Leibenluft, 2011) showed that irritability can be significantly impairing and predictive of anxiety and depression (Evans et al., 2017; Vidal-Ribas et al., 2016). The irritable dimension of ODD is also recognized in the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5; American Psychiatric Association [APA], 2013). The evidence and rationale behind ICD-11's conceptualization of irritability have been detailed by Evans et al. (2017) and Lochman et al. (2015).

ICD-11's interpretation of the evidence base contrasted with DSM-5's. For DSM-5, findings from research on chronic irritability (Brotman et al., 2017; Leibenluft, 2011, 2017) were cited as justification for creating a category named Disruptive Mood Dysregulation Disorder (DMDD) a new childhood depressive disorder (APA, 2013; Roy, Lopes, & Klein, 2014). This line of research began as a response to observed upticks in pediatric BD diagnoses in the United States in the 1990s-2000s, apparently driven some clinicians' interpretation of chronic irritability as a form of pediatric mania. In response, work pioneered by Leibenluft (2011, 2017) and others (for broad reviews, see Brotman et al., 2017; Evans et al., 2017) operationalized chronic irritability as a research syndrome, Severe Mood Dysregulation (SMD),

and generated a compelling body of evidence showing that SMD is distinct from BD. Thus, this research helped clarify the boundaries of pediatric BD; it also demonstrated chronically irritable youth were severely impaired and at increased risk for an array of negative affective, behavioral, and functional outcomes. In consideration of the clear clinical need of this population, SMD was adapted, with modifications, to form DMDD in DSM-5.

As a diagnostic category, however, DMDD lacks compelling evidence for its validity, reliability, and clinical utility (Axelson et al., 2012; Evans et al., 2017; Lochman et al., 2015). Research and clinical advances in this area are sorely needed, but efforts to understand, assess, and treat severe irritability are hindered by inconsistency in how it has been measured and classified. Developing an empirically supported, clinically useful, and globally applicable classification is essential for improving care for youth with severe irritability and related problems. The vast majority of youth with DMDD would already (empirically, if not by definition) receive a diagnosis of ODD, and the DMDD diagnosis shows no incremental validity or utility beyond ODD (Freeman et al., 2016). Clinically, the best available treatment options are not different from those for ODD (Stringaris et al., 2018), but DMDD has raised concerns about the possibility of treatment with psychoactive medications with limited evidence for their efficacy or safety in this population (Tourian et al., 2015).

The revised classification of chronic irritability and DBDDs in ICD-11 must also be evaluated in relation to ICD-10, the previous official system for heath information and reporting and the most widely used classification system globally (Reed et al., 2011). Given that nearly 3 decades have passed since the last major revision, it is not surprising that ICD-11 (WHO, 2020) represents a significant reformulation from ICD-10 (WHO, 1992). For example, ICD-10's F90-F98 section (disorders of childhood and adolescence) has been dissolved, replaced by developmental organization and guidance throughout ICD-11 MBND (Reed et al., 2019). Further, whereas ICD-10 counted ODD as one of several "Conduct disorders," ICD-11 reflects current evidence by recognizing ODD and CD as the two distinct, major categories of DBDDs. Thus, it an important component of the present study was to test whether ICD-11's treatment of DBDDs and irritability, in the context of these other major changes, represents an improvement from ICD-10.

The Present Study

Despite the sound empirical rationale behind these ICD-11 proposals for ODD and irritability, open questions remain regarding their clinical utility and whether they can be used accurately and consistently by clinicians. We sought to investigate these issues in an ICD-11 global field study on DBDDs in youth. In this report, we focus on research questions examining the performance of ICD-11 guidelines as compared to ICD-10 and DSM-5 in the assessment of severe irritability vs. boundary clinical presentations (ODD, bipolar and dysthymic/depressive disorders) and developmentally normative irritability. Specifically, we investigated the extent to which each classification system could assist clinicians in making four key challenging diagnostic distinctions:

1. **Differentiating chronically irritable from non-irritable oppositional behavior.** Most chronically irritable youth, under any diagnostic system, meet the requirements for ODD; yet, many with ODD do not display severe irritability (Evans et al., 2017). This reflects the reality that ODD is heterogeneous, always including disruptive behavior and sometimes including prevailing negative affect dysregulation (Burke & Loeber, 2010). Thus, it is important to test whether clinicians can accurately identify ODD both with and without severe irritability.
2. **Differentiating chronic irritability from episodic BD irritability.** Chronic irritability initially attracted interest because it was purportedly being misdiagnosed as a "broad phenotype" of bipolar disorder (BD)—i.e., some clinicians viewed it as a childhood form of mania (Leibenluft, 2011). If this were the case, certain BD diagnoses (e.g., BD Type II, Other Specified BD) would likely be given. Thus, we investigated this question: Can clinicians differentiate chronic irritability from an episodic BD presentation with irritability?
3. **Differentiating chronic irritability from depressive disorders.** Whereas the preceding question focused on the key distinctions of chronic vs. episodic, it is also important to assess whether clinicians can differentiate among two types of non-episodic, chronic presentations including irritability: ODD vs. chronically depressed mood (i.e., dysthymic/persistent depressive disorder). This distinction is clearly supported by research, with different treatment implications for the two conditions (Stringaris et al., 2018).

4. **Differentiating chronic irritability from normative irritability.** Much of the controversy surrounding DMDD has involved concern about pathologizing developmentally normative irritability, such as temper tantrums in young children (Axelson et al., 2011; Wakefield, 2016). Secondary analyses do not show alarmingly high rates of DMDD, but prevalence varies greatly depending on assessment methods (Axelson et al., 2012; Copeland et al., 2013) and effects on the diagnosis of individual children are unclear. It is possible that including a novel "high severity" disorder may effectively lower the diagnostic threshold for ODD, if the latter is perceived as a "low severity" disorder by comparison.

These Questions 1-4 correspond to Comparisons 1-4 in our experimental design in the Methods, Results, and Discussion that follow. While each specific question is relatively narrow in scope, they collectively served two overarching goals: to both *evaluate* and *refine* the proposed ICD-11 guidelines for better utility in global mental health applications.

METHODS

This project was part of the larger program of ICD-11 field trials for which the broad methods and justification have been published elsewhere (Evans et al., 2015; Keeley et al., 2016). Below we focus only on the design, methodology, and sample for the data reported here.

Participants. A global, multilingual, multidisciplinary sample of mental health professionals (*N*=196) participated in English, Japanese, or Spanish (see Table 1). Clinicians were from 48 countries, with the highest proportions being from India (19.4%), Japan (10.2%), the USA (7.7%), Spain (7.1%), the UK (6.6%), Mexico (5.1%), Nigeria (4.6%), Australia (4.1%), and Canada (4.1%). Participants were identified via the Global Clinical Practice Network (GCPN; https://gcp.network/en/). Beginning in 2011, clinicians from all regions of the globe were invited to register with the GCPN to participate in internet-based field studies for the ICD-11 (Reed et al., 2015). To be eligible for the present study, clinicians had to (a) be a mental health or primary care professional qualified to work with persons with mental disorders; (b) be currently seeing patients or directly supervising services, including (c) at least some services to children; and (d) have proficiency in one of the study languages. Individuals were ineligible to participate if they had contributed to the ICD-11 MBDN revision process or to the DBDD field study

development in any capacity. Informed consent was obtained from all participants. Procedures were exempted from review by the WHO Ethics Review Committee and approved by the Human Subjects Committee of the University of Kansas, where the servers used for data collection were housed.

Development of case vignettes. Five vignettes (labeled A-E; see Table 2 and Supporting Information) were developed to help answer the study questions. Each vignette was designed, refined, and validated to match a target clinical presentation. This work was carried out by DBDD working group members and consultants, following best practices for vignette development (Evans et al., 2015). These experts were asked to write vignettes to meet specified requirements, drawing from actual patient presentations. Draft vignettes were edited for internal consistency, clinical clarity, and cross-cultural applicability, and sent to a separate group of independent experts who conducted confirmatory evaluations of each vignette, to ensure diagnostic agreement for the case narratives. The task of editing and reviewing vignettes involved coding key clinical features (e.g., symptoms, severity, impairment) and identifying any potentially ambiguous aspects, which were then clarified. By the end of this process, each vignette had been vetted to meet the diagnostic requirements for the target disorder under ICD-11, ICD-10, and DSM-5. Any disagreements were resolved by consensus. On this basis, each vignette was designated as having a single "correct" diagnosis per each diagnostic system. Vignettes and other materials were developed in English and translated into Spanish and Japanese using a forward and back-translation process carried out by native speakers with relevant clinical expertise.

Procedures. The study was programmed and hosted in Qualtrics. E-mail invitations were sent to all registered clinicians in each language group who met the participation criteria. When the study was launched, 3,274 GCPN members across all three languages qualified and were contacted via personalized email to invite them to participate. Of these, 686 (21.0% response rate) clicked the link and began the study, of whom 493 (71.9% completion rate) finished the study. The sample of interest here is the subset of 196 participants (39.8% of the total sample) who were allocated to and finished one of the four comparisons (of 13 total comparisons) pertaining to oppositionality and irritability.

Upon following the link and providing consent, clinicians were randomly allocated to the ICD-11, ICD-10, or DSM-5 condition and to a paired comparison condition. Block and stratified

randomization were used to ensure a balance of participant numbers and self-reported DBDD expertise levels across cells. Diagnostic system was masked, although ICD-10 or DSM-5 materials could have been recognized by those familiar with them. After reviewing the diagnostic guidelines, clinicians followed prompts to assess the two assigned vignettes, one at a time (order counterbalanced). To help control for potential gender-related biases, vignette character gender was manipulated such that half the participants assessed two male characters and the other half two female characters (see Supporting Information). The character's age (range 6-11 years) was selected based on clinical-developmental considerations and held constant within vignette. Clinicians could select from broad disorder clusters and specific diagnostic categories, spanning DBDDs, habit/impulse control, bipolar, and depressive disorders. They could review the diagnostic guidelines while deciding and could also enter a different diagnosis or indicate that no diagnosis was warranted. After selecting an initial diagnosis, participants were shown its essential features, one by one, and asked to indicate if each was present in the vignette case. Next, participants could change or re-confirm their diagnosis. These procedures helped ensure that diagnostic decisions were made with due attention to the diagnostic guidelines and to the clinical features of the vignette.

Analytic plan and power. Analyses used a mixed factorial design involving three diagnostic systems (ICD-11, ICD-10, DSM-5) × four paired vignette comparisons (oppositionality, dysthymia, normality; each compared to chronic irritability) as between-participant factors, with a within-participant factor comparing diagnoses of the two paired vignettes. Two-way chi-square statistics were used for bivariate comparisons and the G-square statistic for three-way interactions (Rao & Scott, 1984). Data from all three study languages were combined for analysis. With an average of n=32.1 per pairwise comparison, sensitivity power analyses showed adequate power ($1-\beta$=0.81) to detect large effects (w=0.5) at standard thresholds (α=0.05).

RESULTS

The 196 participating clinicians were randomly allocated to the ICD-11 (n=61), ICD-10 (n=72), and DSM-5 (n=63) conditions, and then to four paired-comparison conditions. The rationale and results for each comparison are described below. See Figure 1 for results, Table 2 for summaries of comparisons and Supporting Information for the full vignettes.

Comparison 1: Can clinicians accurately diagnose irritable vs. non-irritable oppositional behavior?

Comparison 1 evaluated clinicians' assessment of two cases: (1) a child with predominately headstrong/defiant behavior (Vignette B), who would receive ODD under any system; and (2) a child with predominately irritable and angry disruptive behavior (Vignette C), who would be best identified as ODD with Chronic Irritability-Anger under ICD-11, ODD under ICD-10, and DMDD under DSM-5. Participants using ICD-11 selected the correct diagnosis (ODD) for both vignettes 66.7% of the time overall (80.0% and 53.3% respectively), with no difference between vignettes, $\chi^2(1)=2.40$, *ns*. In contrast, participants using both ICD-10 ($\chi^2(1)=7.20$, $p<.01$) and DSM-5 ($\chi^2(1)=13.33$, $p<.001$) were less accurate in their diagnosis of Vignette C compared to Vignette B. Specifically, participants in the ICD-10 condition diagnosed Vignette B with ODD at 77.8% accuracy compared to Vignette C with ODD at 33.3% accuracy. Those using DSM-5 correctly diagnosed Vignette B with ODD at 68.8% accuracy and Vignette C with DMDD with only 6.3% accuracy. The most frequently selected incorrect answers included Unsocialized CD (38.9%) for ICD-10 and ODD (50.0%) for DSM-5. Overall, participants using ICD-11 were more accurate in their diagnostic assignments than those using either ICD-10 ($G^2(4)=10.78$, $p<.05$) or DSM-5 ($G^2(4)=22.78$, $p<.001$).

Comparison 2: Can clinicians accurately diagnose chronic irritability vs. episodic BD irritability?

Participants assigned to this condition viewed two cases: (1) a child with chronic, severe irritability, anger, and outbursts (Vignette C); and (2) a child with irritable/hypomanic mood and depressive episodes, characteristic of BD Type II (Vignette D). Clinicians using ICD-11 ($\chi^2(1)=0.20$, *ns*) and ICD-10 ($\chi^2(1)=0.57$, *ns*) were equally correct for their diagnoses of Vignette C (75.0% and 76.5%, respectively) and Vignette D (66.7% and 64.7%, respectively). Those using DSM-5 showed lower accuracy in assessing Vignette C (4.8%), than Vignette D (38.1%), $\chi^2(1)=6.93$, $p<.01$. Again, those in the DSM-5 condition most often selected ODD (81.0%) for Vignette C although DMDD was correct. Overall, participants in the ICD-11 condition were more accurate than participants using DSM-5 ($G^2(4)=23.78$, $p<.001$), but not different from those using ICD-10 ($G^2(4)=0.80$, *ns*), in identifying episodic bipolar-irritability and chronic oppositional-irritability.

Comparison 3: Can clinicians accurately diagnose chronic irritability vs. dysthymic depression?

Comparison 3 examined this question with two vignettes: (1) a child presenting with chronic irritability and disruptive behavior, characteristic of severe ODD/DMDD (Vignette C); and (2) a child presenting with chronically depressed mood with irritability, representing persistent depression/dysthymia (Vignette E). Participants using ICD-11 were less accurate in diagnosing Vignette E (58.8%) than they were for Vignette C (94.1%), $\chi^2(1)=5.88$, $p<.05$; the specific difficulty was in differentiating among depressive disorders (when all unipolar depressive diagnoses were counted as correct, accuracy was 87.5% and no longer differed from Vignette C, $\chi^2(1)=0.32$, *ns*). Participants using ICD-10 were equally correct for Vignette C (73.7%) and Vignette E (52.6%), $\chi^2(1)=1.81$, *ns*. Those using DSM-5 again had difficulty recognizing DMDD as the correct diagnosis for Vignette C (13.3%) and did better diagnosing Vignette E (60.0%), $\chi^2(1)=7.03$, $p<.01$. The most commonly selected answer for Vignette C for DSM-5 was again ODD (86.7%). Participants in the ICD-11 condition were more accurate overall than those in the ICD-10 ($G^2(4)=9.82$, $p<.05$) and DSM-5 ($G^2(4)=24.54$, $p<.001$) conditions.

Comparison 4: Can clinicians accurately diagnose severe vs. normative irritability?

Two vignettes were presented: (1) a typically developing child with irritable tantrum behavior within normal limits (Vignette A); and (2) a child with a more severe, chronically irritable presentation (Vignette C). Participants using ICD-11 accurately classified Vignette A as normative (88.2%) and Vignette C as clinical (82.4%), with no difference in accuracy between vignettes, $\chi^2(1)=0.23$, *ns*. Unlike previous comparisons, participants using ICD-10 had some difficulty diagnosing Vignette C (33.3%) but this was not worse than their performance on Vignette A (55.6%), $\chi^2(1)=1.80$, *ns*. Vignette C was most often diagnosed as Unsocialized CD (38.9%). Participants using DSM-5 had little accuracy in identifying Vignette A (27.3%) and Vignette C (9.1%), with equally poor accuracy on both, $\chi^2(1)=1.22$, *ns*. In terms of incorrect answers, participants using DSM-5 most often selected ODD for both Vignette A (54.5%) and Vignette C (45.4%). Overall accuracy was greater for participants in the ICD-11 condition compared to those using ICD-10 ($G^2(4)=15.44$, $p<.01$) and DSM-5 ($G^2(4)=28.10$, $p<.001$). Specifically, for typical irritability (Vignette A), those using ICD-11 more often correctly selected “no diagnosis” (88.2%) than those using either ICD-10 (55.6%; $\chi^2(1)=4.58$, $p<.05$) or DSM-5 (27.3%; $\chi^2(1)=10.81$, $p<.01$).

DISCUSSION

We examined the clinical utility of the ICD-11 diagnostic classification of youth irritability and oppositionality, as compared to ICD-10 and DSM-5, among global, multilingual, multidisciplinary clinicians. Participants assigned to use ICD-11 generally made more accurate diagnostic judgments about various presentations of irritability, oppositionality, and mood disturbance as compared to those using the other diagnostic systems. Notably, clinicians using DSM-5 largely failed to assign DMDD when it was appropriate and more often applied psychopathological diagnoses to developmentally typical irritability.

Results suggest that clinicians may struggle to accurately diagnose presentations of chronic irritability using ICD-10 (where answers were scattered across ODD/CD types) and DSM-5 (where they tended to select ODD rather than the correct DSM-5 diagnosis, DMDD). In contrast, ICD-11 may lead to better accuracy overall, and for both irritable and non-irritable presentations of oppositionality. This pattern of results was largely consistent across all comparisons. It may be that when assessing a child with prevailing chronic irritable mood and aggressive temper outbursts, clinicians tend to assign a DBDD diagnosis rather than conceptualize it as a mood (depressive or bipolar) disorder.

Regarding specific boundary presentations, Comparisons 2 and 3 reflect favorably upon clinicians' performance across systems in *identifying* episodic BD with irritability and non-episodic depression with irritability. However, when it comes to *differentiating* these presentations from chronic irritability and oppositionality, those using DSM-5 had more difficulty than those using ICD-11. Although this pattern was most clear and pronounced with DSM-5, ICD-11 also outperformed ICD-10 in Comparison 3, with results suggesting ICD-11 may be especially effective at steering clinicians to the correct overarching category (i.e., differentiating ODD/DBDDs from depressive disorders) as compared to ICD-10 and DSM-5.

Comparison 4 showed that ICD-11 may lead to greater overall accuracy differentiating typical and atypical irritability, compared to both ICD-10 and DSM-5. Within each of these diagnostic systems, participants showed similar levels of accuracy for typical vs. atypical irritability, with ODD being frequently (incorrectly) selected under DSM-5. Notably, clinicians using ICD-11 were more likely to correctly ascertain the *absence* of pathology and indicate that a diagnosis was not warranted. Results

suggest ICD-10 and, especially, DSM-5 could possibly lead to more false positive diagnoses in terms of children receiving *any* diagnosis in response to normative irritable and disruptive behavior.

Study strengths include its global, multilingual, multidisciplinary clinician sample, and well-validated and controlled experimental methodology. Some limitations should be noted. Although vignette-based designs can be valid and reliable research proxies for clinical decision-making (Evans et al., 2015), the use of vignettes remains a limitation. One implication of the vignette approach is that only a select few types of cases can be represented (5 in this study) and contrasted (4 in this study) to investigate central study questions. Our vignettes were not designed to disentangle other important questions, such the diagnostic assessment of youth with severe, explosive temper outbursts but not chronically irritable mood, based on the identification of the phasic vs. tonic components of irritability—an important direction for future work (Brotman et al., 2017; Carlson & Klein, 2018). Research should also consider using actual or simulated patients, either through live clinical interviews or through video-recordings. Additionally, sample size was relatively small per cell; however, power was adequate to detect the large effects observed. Despite the global representation of the sample, it was not possible to test for language or geographical differences in results due to small cell sizes. Moving forward, research evidence generated through a variety of methodologies is needed to better understand and evaluate the ICD-11 formulations of oppositionality and irritability in relation to alternatives. These include epidemiological, clinical (assessment and intervention), behavioral, neuroscientific, and genetic approaches.

Overall, findings support the adequacy and improvement (relative to ICD-10 and DSM-5) of the ICD-11 diagnostic classification and guidance for irritability and oppositionality in youth. Clinicians appear to conceptualize presentations of disruptive mood and behavior problems in accordance with longstanding clinical and scientific understandings of psychopathology, irrespective of which diagnostic guidelines they are asked to use. Indeed, ODD is a well-established heterogenous disorder, which global clinicians showed they can readily diagnose both with and without chronic irritability. This has important clinical implications because effective treatment and management of a condition is predicated on its accurate identification. The literature's recognition that irritability is transdiagnostic and that chronic

irritability predicts *subsequent* internalizing outcomes (Vidal-Ribas et al., 2017) is very useful, and this was part of the basis for DSM-5 locating DMDD within the depressive disorders section. In providing care to a young patient, however, prediction does not trump presentation. Chronic irritability nearly always occurs within ODD; in this context, it can be severe and impairing, warranting clinical attention and predicting homotypic and heterotypic outcomes (Burke & Loeber, 2010; Evans et al., 2017).

Whether chronic irritability is conceptualized as ODD or as DMDD, the best available evidence supports virtually identical treatment recommendations: behavioral parent training and/or cognitive behavior therapy (Kircanski et al., 2018; Stringaris et al., 2018). These well-established psychosocial interventions may serve as first-line treatments, accompanied by a thorough clinical assessment and personalization to address any co-occurring concerns (Evans et al., 2020; Stringaris et al., 2018). This study suggests that ICD-11's coverage of irritability and oppositionality within the DBDD section can be implemented accurately by clinicians and will provide appropriate guidance for clinical decision-making.

KEY POINTS

- The WHO's ICD-11 classified chronic irritability in Oppositional Defiant Disorder (ODD), a formulation differing from ICD-10 and DSM-5.
- In this vignette-based study, we tested the clinical utility of ICD-11's classification of oppositionality and irritability in youth among 196 global clinicians.
- Participants using ICD-11 more accurately identified youths' clinical presentations, while those using DSM-5 made mistakes concerning normative irritability and Disruptive Mood Dysregulation Disorder.
- Results support the clinical utility of ICD-11's ODD formulation and current scientific thinking of ODD as a behavioral disorder that may be accompanied by chronic irritability.
- Careful attention is needed for accurate assessment of youth irritability, with the differential diagnosis likely including ODD, mood disorders, and normative irritability.

Supporting information:

Vignettes Used in the Study (Full Text of Vignettes A-E)

Table S1. Summary of Vignette Target/Correct Diagnoses Per ICD-11, ICD-10, and DSM-5

ACKNOWLEDGMENTS

The authors were members of or consultants to the WHO International Advisory Group for the Revision of ICD-10 Mental and Behavioural Disorders, the Field Studies Coordination Group for ICD-11 Mental and Behavioural Disorders, and/or Subgroup on Disruptive Behaviour and Dissocial Disorders of the Working Group on the Classification of Mental and Behavioural Disorders in Children and Adolescents. GMR was a member of the WHO Secretariat. This article represents the opinions of its authors and except as specifically indicated does not represent the policies or positions of the World Health Organization.

We thank the following individuals for their assistance in developing and testing the vignettes and/or study protocol: Alexandra Monzon, Andrea Garcia, Casey Pederson, Chelsey Hartley, Elena Garralda, Jacky Chan, Jennifer Blossom, Jessy Guler, Jonathan Poquiz, Jürgen Zielasek, Mackenzie Klaver, Salma Siddiqui, and Samantha Burns. For their assistance in translating study materials into Spanish and Japanese, we thank Liz Sosa, Miriam Feria, and Yoko Kamio. The first author gratefully acknowledges support during the preparation of this manuscript from the National Institute of Mental Health and AIM for Mental Health. Lastly, we thank the clinician members of the Global Clinical Practice Network who contributed their valuable time and expertise as participants.

REFERENCES

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. Arlington, VA: American Psychiatric Publishing.

Axelson, D. A., Birmaher, B., Findling, R. L., Fristad, M. A., Kowatch, R. A., Youngstrom, E. A., et al. (2011). Concerns regarding the inclusion of temper dysregulation disorder with dysphoria in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. *Journal of Clinical Psychiatry, 72*, 1257-1262.

Axelson, D., Findling, R. L., Fristad, M. A., Kowatch, R. A., Youngstrom, E. A., Horwitz, S. M., et al. (2012). Examining the proposed disruptive mood dysregulation disorder diagnosis in children in the Longitudinal Assessment of Manic Symptoms study. *Journal of Clinical Psychiatry, 73*, 1342-1350.

Brotman, M. A., Kircanski, K., & Leibenluft, E. (2017). Irritability in children and adolescents. *Annual Review of Clinical Psychology, 13*, 317-341.

Burke, J. D., Boylan, K., Rowe, R., Duku, E., Stepp, S. D., Hipwell, A. E., et al. (2014). Identifying the irritability dimension of ODD: Application of a modified bifactor model across five large community samples of children. *Journal of Abnormal Psychology, 123*, 841-851.

Burke, J., & Loeber, R. (2010). Oppositional defiant disorder and the explanation of the comorbidity between behavioral disorders and depression. *Clinical Psychology: Science and Practice, 17*, 319-326.

Carlson, G. A., & Klein, D. N. (2018). Frying pan to fire? Definition, recognition and treatment challenges of irritability in young people. *Journal of Child Psychology and Psychiatry, 59*, 740-743.

Copeland, W. E., Angold, A., Costello, E. J., & Egger, H. (2013). Prevalence, comorbidity, and correlates of DSM-5 proposed disruptive mood dysregulation disorder. *American Journal of Psychiatry, 170*, 173-179.

Evans, S. C., Burke, J. D., Roberts, M. C., Fite, P. J., Lochman, J. E., de la Peña, F. R., et al. (2017). Irritability in child and adolescent psychopathology: An integrative review for ICD-11. *Clinical Psychology Review, 53,* 29-45.

Evans, S. C., Roberts, M. C., Keeley, J. W., Blossom, J. B., Amaro, C. M., Garcia, A. M., et al. (2015). Vignette methodologies for studying clinicians' decision-making: Validity, utility, and application in ICD-11 field studies. *International Journal of Clinical and Health Psychology, 15,* 160-170.

Evans, S. C., Weisz, J. R., Carvalho, A. C., Garibaldi, P. M., Bearman, S. K., Chorpita, B. F., et al. (2020). Effects of standard and modular psychotherapies in the treatment of youth with severe irritability. *Journal of Consulting and Clinical Psychology, 88,* 255-268.

Ezpeleta, L., Granero, R., de la Osa, N., Penelo, E., & Domènech, J. M. (2012). Dimensions of oppositional defiant disorder in 3-year-old preschoolers. *Journal of Child Psychology and Psychiatry, 53,* 1128-1138.

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.

Freeman, A. J., Youngstrom, E. A., Youngstrom, J. K., & Findling, R. L. (2016). Disruptive mood dysregulation disorder in a community mental health clinic: Prevalence, comorbidity and correlates. *Journal of Child and Adolescent Psychopharmacology, 26,* 12-130.

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.

Keeley, J. W., Reed, G. M., Roberts, M. C., Evans, S. C., Medina-Mora, M. E., Robles, R., et al. (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.

Kircanski, K., Clayton, M. E., Leibenluft, E., & Brotman, M. A. (2018). Psychosocial treatment of irritability in youth. *Current Treatment Options in Psychiatry*, 1-12.

Leibenluft, E. (2011). Severe mood dysregulation, irritability, and the diagnostic boundaries of bipolar disorder in youths. *American Journal of Psychiatry*, *168*, 129-142.

Leibenluft, E. (2017). Irritability in children: What we know and what we need to learn. *World Psychiatry*, *16*, 100-101.

Lochman, J. E., Evans, S. C., Burke, J. D., Roberts, M. C., Fite, P. J., Reed, G. M., et al. (2015). An empirically based alternative to DSM-5's disruptive mood dysregulation disorder for ICD-11. *World Psychiatry*, *14*, 30-33.

Rao, J. N., & Scott, A. J. (1984). On chi-squared tests for multiway contingency tables with cell proportions estimated from survey data. *Annals of Statistics*, *12*, 46-60.

Reed, G. M. (2010). Toward ICD-11: Improving the clinical utility of WHO's International Classification of mental disorders. *Professional Psychology: Research and Practice*, *41*, 457-464.

Reed, G. M., First, M. B., Kogan, C. S., Hyman, S. E., Gureje, O., Gaebel, W., et al. (2019). Innovations and changes in the ICD-11 classification of mental, behavioural and neurodevelopmental disorders. *World Psychiatry*, *18*, 3-19.

Reed, G. M., Rebello, T. J., Pike, K. M., Medina-Mora, M. E., Gureje, O., Zhao, M., et al. (2015). WHO's Global Clinical Practice Network for mental health. *Lancet Psychiatry*, *2*, 379-380.

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

Rowe, R., Costello, E. J., Angold, A., Copeland, W. E., & Maughan, B. (2010). Developmental pathways in oppositional defiant disorder and conduct disorder. *Journal of Abnormal Psychology*, *119*, 726-738.

Roy, A. K., Lopes, V., & Klein, R. G. (2014). Disruptive mood dysregulation disorder: A new diagnostic approach to chronic irritability in youth. *American Journal of Psychiatry*, *171*, 918-924.

Stringaris, A., & Goodman, R. (2009). Three dimensions of oppositionality in youth. *Journal of Child Psychology and Psychiatry*, *50*, 216-223.

Stringaris, A., Vidal-Ribas, P., Brotman, M. A., & Leibenluft, E. (2018). Definition, recognition, and treatment challenges of irritability in young people. *Journal of Child Psychology and Psychiatry, 59,* 721-739.

Tourian, L., LeBoeuf, A., Breton, J.-J., Cohen, D., Gignac, M., Labelle, R., et al. (2015). Treatment options for the cardinal symptoms of disruptive mood dysregulation disorder. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, *24,* 41-54.

Vidal-Ribas, P., Brotman, M. A., Valdivieso, I., Leibenluft, E., & Stringaris, A. (2016). The status of irritability in psychiatry: A conceptual and quantitative review. *Journal of the American Academy of Child & Adolescent Psychiatry*, *55*, 556-570.

Wakefield, J. C. (2016). Diagnostic issues and controversies in DSM-5: Return of the false positives problem. *Annual Review of Clinical Psychology*, *12,* 105-132.

World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines*. Retrieved from: https://www.who.int/classifications/icd/en/bluebook.pdf

World Health Organization. (20192020). *International classification of diseases* (10th and 11th revisions). Retrieved from: https://www.who.int/classifications/icd/en/

Table 1. Participant Characteristics

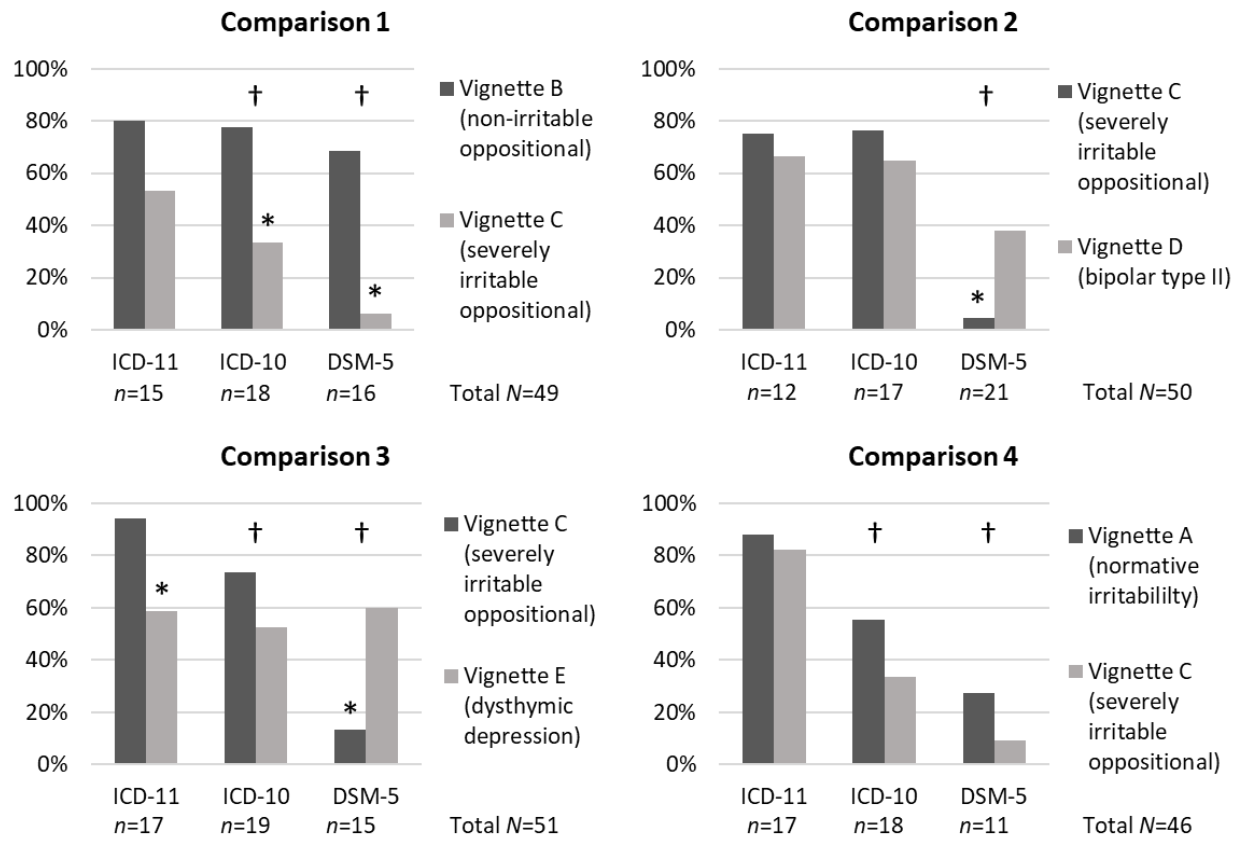
	English	Spanish	Japanese	Total
	f (%)	*f* (%)	*f* (%)	*f* (%)
Region				
AFRO	17 (12.6)			17 (8.7)
AMRO-North	22 (16.3)	1 (2.4)		23 (11.7)
AMRO-South	4 (3.0)	26 (63.4)		30 (15.3)
EMRO	9 (6.7)			9 (4.6)
EURO	32 (23.7)	14 (34.1)		46 (23.5)
SEARO	40 (29.6)			40 (20.4)
WPRO-Asia	2 (1.5)		20 (100)	22 (11.2)
WPRO-Oceania	9 (6.7)			9 (4.6)
Gender				
Male	66 (48.9)	22 (53.7)	15 (75.0)	103 (52.6)
Female	69 (51.1)	19 (46.3)	5 (25.0)	93 (47.4)
Discipline				
Counseling	7 (5.2)			7 (3.6)
Medicine	67 (49.6)	11 (26.8)	16 (80.0)	94 (48.0)
Psychology	51 (37.8)	27 (65.9)	4 (20.0)	82 (41.8)
Social work	4 (3.0)	1 (2.4)		5 (2.6)
Occupational Therapy	3 (2.2)			3 (1.5)
Other	3 (2.2)	2 (4.9)		5 (2.6)
Relevant Specializations*				
Child/Adolescent	62 (45.9)	14 (34.1)	7 (35.0)	83 (42.3)
ADHD/DBDD	41 (30.4)	12 (29.3)	4 (20.0)	57 (29.1)
Total *N* Finished/Analyzed	135	41	20	196
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*
Age	49.83 (11.06)	47.88 (10.89)	50.90 (10.17)	49.53 (10.92)
Years of Experience	17.19 (10.46)	19.17 (9.95)	18.30 (9.67)	17.72 (10.26)

Note. Data were self-reported from GCPN registration. *Specialization variables represent how many clinicians report spending at least 50% of their clinical time working with individuals 0-18 years of age (Child/Adolescent) and how many identified ADHD and Conduct Disorders as one of their top three areas of specialization (ADHD/DBDD).

Table 2. Comparisons, Vignettes, and Correct[a] Diagnoses

	First[b] Vignette	Second[b] Vignette
Comparison 1	*Vignette B* An 8-year-old [boy/girl][c] with predominately headstrong/defiant oppositional behavior. • ICD-11: ODD • ICD-10: ODD • DSM-5: ODD	*Vignette C* A 9-year-old [boy/girl][c] with oppositional behavior and chronic irritability, including mood and outburst features. • ICD-11: ODD • ICD-10: ODD • DSM-5: DMDD
Comparison 2	*Vignette C*	*Vignette D* An 11-year-old [boy/girl][c] with a history of depressive and hypomanic episodes, including irritable features. • ICD-11: Bipolar Type II Disorder • ICD-10: Bipolar affective disorder • DSM-5: Bipolar II Disorder
Comparison 3	*Vignette C*	*Vignette E* A 10-year-old [boy/girl][c] with persistently depressed mood including irritable features. • ICD-11: Dysthymic Disorder • ICD-10: Dysthymia • DSM-5: Persistent Depressive Disorder
Comparison 4	*Vignette A* A 6-year-old [boy/girl][c] with developmentally appropriate tantrums. • ICD-11: no diagnosis • ICD-10: no diagnosis • DSM-5: no diagnosis	*Vignette C*

Note: See Supporting Information for full vignettes. [a]"Correct" diagnoses are those that the vignettes were developed and validated to represent. [b]Order of presentation for a random ~50% of participants (reversed for the other ~50%). [c]A random ~50% of participants within each comparison viewed two male vignettes (the other ~50% viewed two female).



† $p < .05$ for between-condition contrast with ICD-11.

* $p < .05$ for within-condition contrast between the two vignettes.

Figure 1. Percentage of Participating Clinicians Assigning the Correct Diagnosis under ICD-11, ICD-10, and DSM-5 Conditions