



## ORIGINAL ARTICLE

# A case-controlled field study evaluating ICD-11 proposals for relational problems and intimate partner violence



Richard E. Heyman<sup>a,\*</sup>, Cary S. Kogan<sup>b</sup>, Heather M. Foran<sup>c</sup>, Samantha C. Burns<sup>b</sup>, Amy M. Smith Slep<sup>a</sup>, Alexandra K. Wojda<sup>a</sup>, Jared W. Keeley<sup>d</sup>, Tahilia J. Rebello<sup>e</sup>, Geoffrey M. Reed<sup>f</sup>

<sup>a</sup> New York University, USA

<sup>b</sup> University of Ottawa, Canada

<sup>c</sup> University of Klagenfurt, Austria

<sup>d</sup> Virginia Commonwealth University, USA

<sup>e</sup> Columbia University, USA

<sup>f</sup> World Health Organization, Switzerland

Received 1 December 2017; accepted 8 March 2018

Available online 7 April 2018

### KEYWORDS

International Classification of Diseases;  
Intimate partner violence;  
Intimate partner relationship problems;  
Mental health problems;  
Case-controlled field study

**Abstract** *Background/Objective:* Intimate partner relationship problems and intimate partner abuse and neglect — referred to in this paper as “relational problems and maltreatment” — have substantial and well-documented impact on both physical and mental health. However, classification guidelines, such as those found in the International Classification of Diseases (ICD-10), are vague and unlikely to support consistent application. Revised guidelines proposed for ICD-11 are much more operationalized. We used standardized clinical vignette conditions with an international panel of clinicians to test if ICD-11 changes resulted in improved classification accuracy. *Method:* English-speaking mental health professionals ( $N = 738$ ) from 65 nations applied ICD-10 or ICD-11 (proposed) guidelines with experimentally manipulated case presentations of presence or absence of (a) individual mental health diagnoses and (b) relational problems or maltreatment. *Results:* ICD-11, compared with ICD-10, guidelines resulted in significantly better classification accuracy, although only in the presence of co-morbid mental health problems. Clinician factors (e. g., gender, language, world region) largely did not impact classification performance. *Conclusions:* Despite being considerably more explicated,

\* Corresponding author: Family Translational Research Group, New York University, 137 East 25th Street #603, New York, NY 10010, USA, 011-212-998-9984, USA.

E-mail address: [Richard.heyman@nyu.edu](mailto:Richard.heyman@nyu.edu) (R.E. Heyman).

**PALABRAS CLAVE**

Clasificación Internacional de Enfermedades; violencia de pareja; problemas en la relación de pareja; problemas de salud mental; estudio de campo con casos controlados

raters' performance with ICD-11 guidelines reveals training issues that should be addressed prior to the release of ICD-11 in 2018 (e. g., overriding the guidelines with pre-existing archetypes for relationship problems and physical and psychological abuse).

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

### Estudio de campo con casos controlados para evaluar propuestas de la CIE-11 en problemas relacionales y de violencia de pareja

**Resumen** *Antecedentes/Objetivo:* Los problemas en la relación de pareja y relacionados con abuso y negligencia de pareja, referidos como "problemas relacionales y maltrato", tienen un importante impacto en la salud física y mental. Sin embargo, guías de clasificación, como la Clasificación Internacional de Enfermedades (CIE-10), son vagas y su aplicación es inconsistente. Las guías propuestas por el CIE-11 son más operacionales. Junto con un panel de clínicos, utilizamos viñetas clínicas estandarizadas, para evaluar si los cambios propuestos por CIE-11 mejoraban la precisión de la clasificación. *Método:* Profesionales de la salud de habla inglesa ( $N=738$ ) de 65 naciones compararon la aplicación del CIE-10 y CIE-11 en casos experimentales, estableciendo presencia o ausencia de (a) diagnósticos individuales de salud mental y (b) problemas de relaciones o maltrato. *Resultados:* CIE-11 tuvo resultados significativamente más precisos, aunque solo en presencia de comorbilidades de salud mental. Factores como género, idioma y región no presentaron mayor alteración. *Conclusiones:* Aunque el CIE-11 está mejor explicado, este estudio revela problemas de capacitación que deberían abordarse antes de su publicación en 2018.

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

The health impacts of intimate partner relationship problems (Kiecolt-Glaser & Wilson, 2017; Robles, Slatcher, Trombello, & McGinn, 2014), intimate partner maltreatment (i. e., partner physical, emotional, and/or sexual abuse and partner neglect; Coker et al., 2002; Lagdon, Armour, & Stringer, 2014) have been well documented. Grouped here as "Relational Problems and Maltreatment" (RPMs), each of these problems has extensive research literatures on prevalence, etiology, and treatment (Sullivan & Lawrence, 2016; Bray & Stanton, 2012; Foran, Beach, Slep, Heyman, & Wamboldt, 2013); are among the most common themes in psychotherapy (Gaut, Steyvers, Imel, Atkins, & Smyth, 2017); and are factors in precipitating, exacerbating, and maintaining mental and behavioral disorders (Schonbrun & Whisman, 2010).

In recognition of the importance of couple and family health to worldwide physical and mental health, the WHO International Advisory Group for the Revision of the International Classification of Diseases (ICD) Mental and Behavioral Disorders created a Working Group to develop evidence-based proposals for improving the usability of the ICD's definitions that assist clinicians in reliably identifying RPMs. The Working Group noted that (a) the ICD-10 RPM guidelines are vague and unlikely to support consistent application; and (b) RPMs are found in multiple places in the ICD-10 (e. g., Z, T, Y codes). The Working Group recommended that these factors be consolidated and revised to enhance clinical utility (i. e., the ability of a classification system

to facilitate communication among stakeholders; support implementation and useful clinical management across clinical settings; and facilitate improvements in individual- and population-level health outcomes; Reed, 2010; Reed et al., 2013).

This study is part of a program of developmental field studies WHO is conducting to inform the ICD revision, expected to be available in 2018. These studies use clinical vignettes to evaluate experimentally the impact of proposed changes to the ICD definitions for mental and behavioral disorders on clinician diagnostic behavior (Reed et al., 2013). Because the ICD provides a global classification of all health conditions and a shared nomenclature for clinicians worldwide, an important element of clinical utility evaluation is testing the proposed guidelines with users from myriad national, lingual, and disciplinary backgrounds (Reed, 2010).

The Working Group's proposals for ICD-11 RPM definitional requirements for maltreatment were adapted from Heyman and Slep's (2006) criteria, which were independently developed and field-tested in a prior five-study program that included a content validity study, a mixed-method study with clinicians about clinical utility, development of operationalized criteria, evaluation of the inter-rater agreement of the revised criteria under typical usage in field settings, and evaluation of the inter-rater agreement of the revised criteria using a computerized decision support tool. Baseline agreement between field-

users at 5 sites and master reviewers was 50%; in the final development field trial, agreement was 92% (Heyman & Slep, 2006) and was maintained at 91% when the criteria were disseminated to 41-sites worldwide (Heyman & Slep, 2009). Evidence of content and criterion validity of the maltreatment criteria have been documented across multiple studies (Heyman & Slep, *in press*). The criteria for intimate partner relationship problems and parent-child relationship problems, although subject to less extensive developmental studies, also have promising inter-rater agreement and validity research supportive of further testing (Heyman & Slep, *in press*).

The Working Group noted that although the ICD is used in worldwide health settings, the past research was conducted in specific and circumscribed health settings (i. e., U. S. military health agencies [maltreatment criteria] and U. S. academic health settings [partner relationship problem criteria]). Although prior research documented U. S. content validity (Heyman & Slep, 2006) and clinical utility (Heyman, Collins, Slep, & Knickerbocker, 2010) (further bolstered by their adaptation for the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5; American Psychiatric Association, 2013), ICD pre-adoption research has specific, usage-related research questions that require attention. For instance, the U.S. family protective services settings use the maltreatment RPM criteria guidelines in isolation, whereas clinicians employing the ICD internationally will use them in combination with other health diagnoses including other mental and behavioral disorder diagnoses. Furthermore, raters in prior studies and settings have been required to make criterion-by-criterion decisions using either a computerized decision support tool or structured clinical interview, whereas clinicians employing the ICD use definitional requirements, which are the minimally required features needed to make presence/absence decisions.

This paper focuses on the proposed ICD-11 intimate partner RPMs. Our goals were, first, to compare the classification accuracy of an international sample of mental health professionals using Z-codes (i. e., "Factors influencing health status and contact with health services") from ICD-10 and the proposed ICD-11 guidelines across four vignette conditions with experimentally manipulated case presentations of (a) presence or absence of individual mental health diagnoses (i. e., Major Depressive Disorder [MDD], General Anxiety Disorder [GAD]) and (b) presence or absence of RPMs; and, second, to test the extent that clinician-related, rather than clinical case presentation, factors related to variation in classification performance. Vignette methodology is uniquely suited to test the specific effects of the different guidelines on diagnostic decision-making (Keeley et al., 2016) and responses of clinicians to vignettes tend to be generalizable to decision-making in "real world" clinical settings (Evans et al., 2015).

We hypothesized that the more specified ICD-11 guidelines would outperform the ICD-10 guidelines. We investigated whether real world moderators, such as clinician characteristics or presence of mental health diagnoses, impacted performance, but did not have a priori hypotheses.

## Method

### Participants

Participants were drawn from the Global Clinical Practice Network (GCPN), a worldwide network of over 14,000 mental health professionals formed for the ICD-11 case-controlled field trials. Members of the GCPN were recruited through international and national conferences in psychology, psychiatry and related disciplines; national and regional professional associations; professional listservs; and by word of mouth. Although not strictly representative of clinicians worldwide, GCPN membership is large and diverse, arguing for generalizability of results and allowing for meaningful comparisons of clinician characteristics (Keeley et al., 2016). See Reed et al. (2015) for more information on the history and development of the GCPN.

The study was conducted in English. The sampling frame was GCPN clinicians whose registration indicated both (a) self-reported proficiency or fluency in English, and (b) current provision of mental health clinical services with patients or engagement in direct clinical supervision. GCPN members ( $N = 5,686$ ) meeting these criteria were sent a personalized email invitation. Reminder emails were sent two and four weeks later, and data collection lasted two months; 1,421 (25%) responded to the survey link and began the study; 75 participants who reported not meeting eligibility criteria were eliminated. Of the remaining 1,346, 738 (55%; 13% of total invited) from 65 nations completed the study and their data were used for analysis. Compared with clinicians who were invited but did not participate, completers did not differ significantly on age, years of experience, and gender but were slightly (a) more likely to be social workers and less likely to be certified peer support workers and (b) more likely to come from the African region and less likely to come from North America. Descriptive participation information can be found in Table 1.

### Materials

Materials for the study included the proposed Relationship Problem and Maltreatment (RPM) codes and features for ICD-11, RPM codes (i. e., Z-codes) for ICD-10, and proposed Mental and Behavioral Disorder (MBD) diagnostic definitions for ICD-11 (i. e., Mood Disorders and Anxiety and Fear-Related Disorders).

Vignettes were developed and tested according to standard procedures for ICD-11 field trials (e. g., Evans et al., 2015; Keeley, Reed, Roberts, Evans, Robles et al., 2016). Twelve case vignettes were generated for the study (see Table 2). Drs. Heyman, Slep, and Foran drafted the vignettes — based on actual clinical cases — to contain both key characteristics (i. e., MBD symptoms and RPM codes) and typical clinical presentations; "gold standard answers" for the vignettes were subsequently validated by 12 international RPM experts to ensure consensus agreement on the correct codes and diagnoses, and global applicability of the cases. Vignettes with less than 90% agreement across raters were revised by Drs. Reed and Kogan to address identified areas of disagreement. Vignettes described male and

**Table 1** Demographic backgrounds of participating clinicians.

WHO Global Region	<i>n</i> (%)
<i>Africa</i>	33 (4.5)
<i>Americas, North and South</i>	203 (27.5)
<i>Eastern Mediterranean</i>	34 (4.6)
<i>Europe</i>	295 (40.0)
<i>South-East Asia</i>	107 (14.5)
<i>Western Pacific, Asia and Oceania</i>	66 (8.9)
<i>Gender</i>	
Male	384 (52.0)
Female	353 (47.9)
Other	1 (0.1)
<i>Discipline</i>	
Certified Peer Support Work <sup>a</sup>	5 (0.7)
Counseling	38 (5.1)
Medicine	341 (46.2)
Nursing	17 (2.3)
Occupational Therapy	30 (4.1)
Psychology	250 (35.9)
Social Work	35 (4.7)
Sex Therapy	4 (0.5)
Speech Therapy	1 (0.1)
Other	18 (2.4)
	<i>M (SD)</i>
Age	48.40 (11.41)
Years of experience	16.31 (10.40)

<sup>a</sup> Certified Peer Support Workers are individuals with lived experience that obtain certification to provide support services to patients. Certification exists in several countries around the globe (Canada, UK, US, New Zealand, Australia and others). CPSW do not generally receive diagnostic training nor do they have diagnostic privileges. See [Jacobson, Trojanowski, and Dewa \(2012\)](#) for additional information on services provided.

female adults across a range of ages who were currently in heterosexual relationships but omitted specific detail associated with any cultural group or religious practice. The vignettes reflected the four study conditions: (a) features consistent with both a RPM and a MBD, (b) features consistent with only a RPM (c) features consistent with only a MBD and (d) features consistent with neither a RPM nor a MBD (see [Table 2](#)). Specifically, each vignette described an individual or couple experiencing either the presence or absence of one of three ICD-11 RPMs (i. e., Relationship Distress with Spouse or Intimate Partner; Spouse or Partner Violence, Physical; or Spouse or Partner Abuse, Psychological) and the presence or absence of one of two common MBDs (i. e., Single Episode Depressive Disorder or Generalized Anxiety Disorder). Vignettes in which a RPM is absent described subthreshold features for the RPM, and vignettes in which a MBD is absent described subthreshold mental health symptoms. Due to concerns about the length of the study (i. e., time to complete), two lower prevalence ICD-11 RPM categories — (1) Spouse or Partner Violence, Sexual and (2) Spouse or Partner Neglect — were not included.

## Procedure

This study was approved by the Human Subjects Committee at the University of Kansas, Lawrence Campus (HSCL #20804) and exempted from review by the World Health Organization Research Ethics Review Committee (Protocol ID RPC569). Participants received an email invitation to participate in the study and were asked to follow an individualized link to the survey in Qualtrics. Upon entry to the study, participants were randomly assigned to view either ICD-10 or ICD-11 RPM Z-codes and their corresponding features without an explicit statement about which classification they would use in the remainder of the experiment. Participants in both conditions had access to ICD-11 MBD diagnostic definitions to assist with determining whether individuals described in the vignettes exhibited the essential features of a mood or anxiety and fear-related disorder. These definitions were included because assessing clinicians typically do not face choices among RPMs only (or no diagnosis); symptoms are often discussed in a jumble of individual and relational symptoms and contexts. Thus, this study's vignettes provided an analogue of clinicians making diagnostic decisions in the context of either individual or relational diagnoses (or both or neither). However, empirical evaluation of the individual disorder requirements was not an aim of the present study.

Once participants reviewed either ICD-10 or ICD-11 codes and ICD-11 MBD diagnostic definitions, they were further randomly assigned to one of six comparison conditions (see [Table 3](#)). Comparisons were devised according to a logical semi-random assignment process, where each vignette was equally represented throughout the study and had a similar probability of being presented with each of the other vignettes. Within each comparison, participants were presented with four vignettes, each selected from a different study condition (see [Table 3](#)). Within the four vignettes, the same RPM code was presented a maximum of two times (e. g., presence of Spouse or Partner Violence, Physical, and absence of Spouse or Partner Violence, Physical). Within MBDs, depression or GAD could appear a minimum of zero and a maximum of four times. The order of presentation of the four vignettes was counterbalanced across participants.

After reading each vignette, participants were asked to provide a MBD diagnosis, followed by a RPM code from lists of MBD and RPM categories (ICD-10 RPM or ICD-11 RPM Z codes and ICD-11 MBD categories). Participants were able to refer to MBD diagnostic definitions and descriptions of RPM codes when making their decisions. They then rated the presence or absence of each of the essential features of the assigned RPM code (ICD-10 or ICD-11) for the specific code they selected. The rationale for the individual rating was that all features must be met for a RPM to be correctly assigned. Therefore, this component of the study provides the participant an opportunity to explicitly evaluate whether the RPM they are assigning met the definition according to ICD. After reviewing each essential feature for the RPM code, participants were given the option of changing their selected code and their selected mental and behavioral disorder diagnosis.

If the selected final RPM code was incorrect, participants were asked to indicate in narrative form why they

**Table 2** Correct responses for vignettes according to ICD-11 and ICD-10.

Condition	Vignettes	Correct ICD-11 RPM code	Correct ICD-10 RPM code	Correct MBD ICD-11 diagnosis
I (RPM present, MBD present)	A	Relationship Distress with Spouse or Intimate Partner	Problems in relationship with spouse or partner	Single Episode Depressive Disorder
	C	Spouse or Partner Violence, Physical	Problems in relationship with spouse or partner	Single Episode Depressive Disorder
	E	Spouse or Partner Abuse, Psychological	Problems in relationship with spouse or partner	Generalized Anxiety Disorder
II (RPM present, MBD absent)	B	Relationship Distress with Spouse or Intimate Partner	Problems in relationship with spouse or partner	None (subthreshold depression)
	D	Spouse or Partner Violence, Physical	Problems in relationship with spouse or partner	None (subthreshold depression)
	F	Spouse or Partner Abuse, Psychological	Problems in relationship with spouse or partner	None (subthreshold depression)
III (RPM absent, MBD present)	G	None (subthreshold Relationship Distress with Spouse or Intimate Partner)	None	Generalized Anxiety Disorder
	I	None (subthreshold Spouse or Partner Violence, Physical)	None	Single Episode Depressive Disorder
	K	None (subthreshold Spouse or Partner Abuse, Psychological)	None	Generalized Anxiety Disorder
IV (RPM absent, MBD absent)	H	None (subthreshold Relationship Distress with Spouse or Intimate Partner)	None	None (subthreshold Generalized Anxiety Disorder)
	J	None (subthreshold Spouse or Partner Violence, Physical)	None	None (subthreshold depression)
	L	None (subthreshold Spouse or Partner Abuse, Psychological)	None	None (subthreshold depression)

*Note.* ICD: International Classification of Diseases; MBD: Mental and Behavioural Disorders; RPM: Relationship Problems and Maltreatment. Each vignette includes referral information, presenting problems and additional background information.

**Table 3** Study comparisons.

Comparison	Condition I (MH+ RPM+)	Condition II (MH- RPM+)	Condition III (MH+ RPM-)	Condition IV (MH- RPM-)
1	A	F	K	H
2	A	D	I	L
3	C	B	G	L
4	C	F	K	J
5	E	D	I	H
6	E	B	G	J

*Note.* Condition I = RPM present, mental and behavioral disorder present; Condition II = RPM present, mental and behavioral disorder absent; Condition III = RPM absent, mental and behavioral disorder present; Condition IV = RPM absent, mental and behavioral disorder absent.

had assigned the selected RPM code rather than the correct response (without explicitly identifying the selected RPM code as incorrect). Participants then completed the sequence again for a second, third, and fourth vignette. Finally, participants were asked to rate their level of familiarity with RPM codes from different diagnostic manuals (i. e., Z, T or Y-codes in ICD-10, V codes in DSM-5 or DSM-IV [American Psychiatric Association, 1994]).

## Results

Because ICD-10 Z-codes contain only an omnibus “Problems in relationship with spouse or partner,” comparisons with ICD-11 necessarily required collapsing ICD-11 respondents’ classifications into “yes/no” for all four RPM types tested. However, subsequent analyses for individual ICD-11 RPM types were used to investigate

**Table 4** Accuracy of clinicians' classifications using ICD-11 (Proposed) and ICD-10 Criteria.

ICD Criteria	Condition I (MH+ RPM+)		Condition II (MH- RPM+)		Condition III (MH+ RPM-)		Condition IV (MH- RPM-)	
	MH	RPM	MH	RPM	MH	RPM	MH	RPM
ICD-11 (proposed)	72%	84%	80%	89%	78%	73%	78%	64%
ICD-10	67%	74%	78%	84%	78%	62%	73%	62%
$\chi^2(1)$	1.60	10.83 <sup>†</sup>	0.43	4.51	0.001	10.34 <sup>†</sup>	2.24	0.39
Cramer's V	0.047	0.121	0.024	0.078	0.001	0.118	0.055	0.023

Note. Classification accuracy could range between 0% (no agreement with the correct vignette classification) to 100% (complete agreement with the correct vignette classification). <sup>†</sup>  $p < .05$  with Benjamini-Hochberg False Discovery Rate correction. Note: ICD = International Classification of Diseases; MH = Mental Health Problems; RPM = Relationship Problems and Maltreatment; + = Present; - = Absent;  $N = 738$ .

the accuracy of clinician decisions using the proposed guidelines.

### Overall performance across MH and RPM

Overall, clinicians using the proposed ICD-11 ( $M = 0.78$ ,  $SD = 0.20$ ), compared with those using ICD-10 ( $M = 0.70$ ,  $SD = 0.20$ ), guidelines were more likely to correctly apply RPM classifications ( $t(736) = 4.845$ ,  $p < .001$ ,  $d = 0.357$ ); not surprisingly, having access to proposed ICD-11 ( $M = 0.77$ ,  $SD = 0.24$ ) or ICD-10 ( $M = 0.74$ ,  $SD = 0.25$ ) RPM guidelines did not notably affect clinicians' overall performance for MH disorders ( $t(736) = 1.52$ ,  $p = .127$ ,  $d = 0.11$ ).

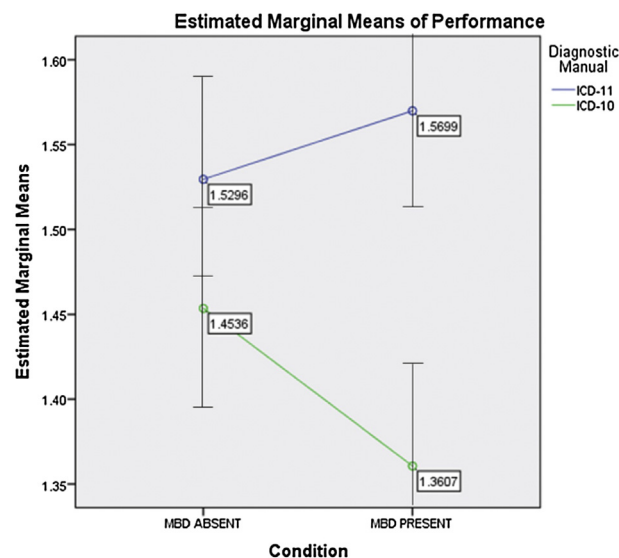
### Performance by condition/vignette

Table 4 shows clinicians' performance for any RPM (presence/absence) x MH (presence/absence) combination. Chi-square analyses were conducted for each ICD-10 and ICD-11 comparison (i. e., the frequency of correct responses for the ICD-10 versus the ICD-11 for RPM or for MH). For clinicians using ICD-11, compared with those using ICD-10, RPM guidelines were significantly more likely to correctly classify RPM presence/absence when a MH condition was present (see results for Conditions I and III in Table 4). When a MH condition was absent, there were no significant difference between clinicians using proposed ICD-11 and ICD-10 guidelines (although the direction of non-significant difference always favored ICD-11). A mixed measures ANOVA was run to test this finding across conditions. There was a significant interaction between condition (MH present I and III versus MH absent II and IV) and ICD version [ $F(1, 736) = 4.502$ ,  $p = .034$ ;  $\eta^2 = .006$ ]. As shown in Figure 1, ICD-11 outperformed ICD-10, with more pronounced differences when MH conditions were absent.

### Differences by clinician characteristics

First, using GCPN clinician registration information (see Reed et al., 2015), we tested whether gender, language, region, profession, age, and years of clinical experience impacted classification accuracy for ICD-11 (proposed) and ICD-10 guidelines. There were no significant interactions for any demographic factor.

Next, we tested the impact of demographic factors on diagnostic accuracy when responses were collapsed across



**Figure 1** Estimated marginal means of classification accuracy for ICD-11 vs. ICD-10 by condition.

ICD-10 and ICD-11 conditions (see Table 5). To control for family-wise error, analyses were conducted for each condition applying the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995). Significant differences were found for age and world region. Age affected judgments when MH problems were present and RPMs were absent, with both older (61 and older) and younger (20–30 years old) misclassifying the MH problems as not being present compared with the middle age groups (the three groups between 31–60 years old). Region impacted performance in one condition: when a MH problem was absent but an RPM was present, clinicians from Africa or South East Asia (67–68% accuracy) were more likely to misclassify by scoring the MH problem as being present compared with clinicians from Europe (85% accuracy).

### ICD-11 (Proposed) Performance for Specific RPMs

Unlike ICD-10 Z-codes, proposed ICD-11 Z-codes differentiate among RPMs; accuracy for specific problems is reported in Table 6. Accuracy was highest for "relationship distress with a spouse or intimate partner" (82%–89%) and lower for partner physical or psychological abuse (45%–78%). Table 7 shows the classification/misclassification patterns. For both

**Table 5** Classification accuracy by demographic variables by study condition (chi-square and exact tests).

Demographic Factor	Condition I (MH + RC +)		Condition II (MH- RPM)		Condition III (MH + RPM)		Condition IV (MH- RPM-)	
	MH	RPM	MH	RPM	MH	RPM	MH	RPM
<i>Gender</i> ( $\chi^2$ )	8.96	3.66	1.74	2.03	3.75	2.14	3.4	1.22
Male (n= 384)	65%	64%	77%	75%	77%	68%	73%	61%
Female (n=353)	75%	69%	81%	70%	79%	68%	79%	64%
Cramer's V	0.11	0.016	0.049	0.052	0.071	0.054	0.068	0.041
<i>Age</i> ( $\chi^2$ )	2.27	6.15	5.99	9.29	17.81 <sup>*</sup>	9.11	6.94	8.14
20-30 (n=57)	67%	75%	79%	65%	63% <sup>a</sup>	79%	81%	53%
31-40 (n=196)	69%	61%	81%	77%	81% <sup>b</sup>	67%	74%	66%
41-50 (n=229)	70%	67%	81%	68%	82% <sup>b</sup>	69%	74%	63%
51-60 (n=175)	72%	67%	79%	77%	79% <sup>b</sup>	69%	80%	64%
61-70 (n=65)	63%	72%	71%	75%	69% <sup>a</sup>	57%	74%	51%
71+ (n=16)	75%	63%	63%	56%	56% <sup>a</sup>	63%	63%	69%
Cramer's V	0.055	0.091	0.09	0.112	0.155	0.1	0.082	0.103
<i>Years experience</i> ( $\chi^2$ )	5.62	3.16	3.34	6.56	3.39	0.51	4.55	4.21
0-5 (n=314)	70%	64%	82%	74%	80%	69%	75%	60%
6-10 (n=226)	65%	66%	78%	66%	77%	68%	74%	66%
11-20 (n=148)	75%	72%	76%	77%	78%	68%	80%	64%
21-30 (n=41)	71%	66%	76%	76%	76%	63%	78%	54%
31-40 (n=9)	56%	78%	67%	78%	56%	67%	56%	67%
Cramer's V	0.087	0.065	0.067	0.094	0.068	0.026	0.078	0.075
<i>Discipline</i> ( $\chi^2$ )	1.52	7.93	11.39	7.07	15.64	17.79	12.96	4.16
Medicine (n=341)	70%	63%	78%	75%	81%	72%	71%	64%
Psychology (n=250)	68%	68%	83%	67%	80%	68%	80%	62%
Nursing (n=17)	71%	65%	88%	71%	59%	65%	88%	47%
Other (n=27)	67%	78%	63%	74%	59%	52%	74%	56%
Counseling (n=38)	74%	71%	66%	79%	68%	58%	71%	55%
Occupational therapy (n=30)	77%	73%	77%	77%	67%	80%	90%	70%
Social work (n=35)	66%	77%	80%	80%	77%	46%	77%	63%
Cramer's V	0.045	0.108	0.124	0.098	0.146	0.155	0.133	0.075
<i>Regional language</i> ( $\chi^2$ )	4.64	10.9	6.75	3.23	1.5	7.4	6.89	7.89
Chinese (n=12)	58%	25%	67%	75%	67%	92%	83%	75%
English (n=655)	70%	67%	80%	73%	78%	67%	76%	62%
French (n=32)	59%	69%	69%	75%	81%	81%	63%	56%
German (n=24)	75%	63%	71%	79%	79%	71%	92%	88%
Portuguese (n=10)	50%	50%	60%	50%	70%	50%	70%	60%
Cramer's V	0.08	0.122	0.096	0.066	0.045	0.1	0.097	0.104
<i>Region</i> ( $\chi^2$ )	11.99	10.52	19.25 <sup>*</sup>	0.89	11.09	5.826	1.39	3.543
Africa	79%	88%	67% <sup>a</sup>	85%	70%	70%	73%	67%
Americas, North and South	71%	85%	76% <sup>ab</sup>	87%	76%	61%	76%	62%
Eastern Mediterranean	62%	77%	82% <sup>ab</sup>	88%	79%	65%	71%	53%
Europe	69%	75%	85% <sup>b</sup>	85%	83%	71%	77%	65%
South-East Asia	60%	82%	68% <sup>a</sup>	88%	70%	71%	73%	59%
Western Pacific, Asia and Oceania	82%	73%	80% <sup>ab</sup>	86%	74%	68%	76%	67%
Cramer's V	0.127	0.119	0.162	0.035	0.123	0.089	0.043	0.069
<i>Language proficiency: ICD-11</i> ( $\chi^2$ )	0.003	3.72	0.05	1.21	0.98	0	1.39	1.98
ICD-11: Completely fluent (n=115)	71%	57%	79%	71%	75%	73%	82%	69%
ICD-11: Advanced (n=257)	72%	67%	80%	77%	79%	73%	76%	61%
Cramer's V	0.003	0.1	0.012	0.057	0.051	0.001	0.061	0.073
<i>Language proficiency: ICD-10</i> ( $\chi^2$ )	2.87	0.26	2.17	1.09	0.04	0.43	0.37	1.13
ICD-10: Completely fluent (n=107)	61%	71%	73%	74%	79%	65%	71%	65%
ICD-10: Advanced (n=259)	70%	68%	80%	68%	78%	62%	74%	60%
Cramer's V	0.089	0.026	0.077	0.054	0.01	0.034	0.032	0.056

**Table 5** (Continued)

Demographic Factor	Condition I (MH + RC +)		Condition II (MH- RPM)		Condition III (MH + RPM)		Condition IV (MH- RPM-)	
	MH	RPM	MH	RPM	MH	RPM	MH	RPM
<i>Familiarity with codes: ICD-11 (<math>\chi^2</math>)</i>	7.81	0.21	1.16	0.67	0.66	0.003	2.79	0.05
ICD-11: Use once per week or more	75%	63%	81%	76%	79%	73%	80%	64%
ICD-11: Use less often	60%	66%	76%	72%	75%	73%	72%	63%
Cramer's V	0.145	0.023	0.056	0.043	0.042	0.003	0.087	0.012
<i>Frequency of ICD use: ICD-10 (<math>\chi^2</math>)</i>	2.65	4.12	1.73	0.18	2.82	5.854	3.29	0.33
ICD-10: Use once per week or more	70%	66%	80%	69%	81%	67%	76%	62%
ICD-10: Use less often	62%	76%	74%	71%	73%	54%	67%	59%
Cramer's V	0.085	0.106	0.069	0.022	0.088	0.127	0.095	0.03

Note. Classification accuracy could range between 0% (no agreement with the correct vignette classification) to 100% (complete agreement with the correct vignette classification). \* Significant according to Benjamini-Hochberg correction. Means sharing the same superscript are not significantly different from each other.

**Table 6** ICD-11 (proposed) accuracy, by RPM and study condition (where RPM was present).

	Condition I (MH + RPM +)	Condition II (MH- RPM +)
Relationship distress with spouse or intimate partner	89% <sup>a</sup>	82% <sup>a</sup>
Spouse or partner violence, physical	45% <sup>b</sup>	65% <sup>b</sup>
Spouse or partner abuse, psychological	59% <sup>b</sup>	78% <sup>ab</sup>
$\chi^2(2)$	50.72 <sup>*</sup>	11.81 <sup>*</sup>
Cramer's V	0.370	0.178

Note. Classification accuracy could range between 0% (no agreement with the correct vignette classification) to 100% (complete agreement with the correct vignette classification). \*  $p < .05$  with Benjamini-Hochberg False Discovery Rate correction. Means sharing the same superscript are not significantly different from each other. Note: MH = Mental Health Problems; RPM = Relationship Problems and Maltreatment; + = Present; - = Absent.

**Table 7** Clinician responses by Condition (ICD-11 only).

Correct response	Relationship distress with spouse or intimate partner	Spouse or partner violence, physical	Spouse or partner abuse, psychological	Spouse or partner violence, sexual	Spouse or partner neglect	No RPM
<i>Condition I; MH+RPM+</i>						
Relationship distress with spouse or intimate partner	<b>89%</b>	0%	4%	0%	2%	5%
Spouse or partner violence, physical	18%	<b>45%</b>	4%	1%	0%	33%
Spouse or partner abuse, psychological	28%	4%	<b>59%</b>	0%	0%	9%
<i>Condition II; MH-RPM+</i>						
Relationship distress with spouse or intimate partner	<b>84%</b>	0%	5%	0%	1%	10%
Spouse or partner violence, physical	15%	<b>65%</b>	2%	0%	0%	18%
Spouse or partner abuse, psychological	18%	0%	<b>78%</b>	0%	0%	4%

Note. Classification accuracy could range between 0% (no agreement with the correct vignette classification) to 100% (complete agreement with the correct vignette classification). **Bold** indicates correct classification. MH = Mental Health Problems; RPM = Relationship Problems and Maltreatment; + = Present; - = Absent.



psychological and physical abuse, clinicians commonly misclassified abuse as “relationship distress with a spouse or intimate partner” or, somewhat disconcertingly, as no RPM at all. After completing the vignettes, the clinicians were queried about their responses in a open-ended web form. Most of the misclassifications appeared to come from clinicians substituting their own implicit abuse criteria for those in the ICD materials (e.g., believing that acts of “abuse” would need to be chronic or pervasive to be classified as such, and otherwise defaulting to “relationship distress”).

## Discussion

This is the first study of the classification accuracy of ICD RPM codes. By experimentally manipulating case presentations, we were able to test clinicians’ diagnostic accuracy using either proposed ICD-11 descriptors for three adult intimate partner RPMs (relationship distress; physical violence, psychological abuse) or the existing ICD-10 single category (problems in relationship with spouse/partner) when combined with the presence or absence of individual mental health disorders. Prior clinic-based field testing of the guidelines (Heyman & Slep, 2006, 2009) indicated that the greater specificity of the guidelines for RPMs led to substantial increases in diagnostic accuracy. Thus, we expected the more thoroughly delineated ICD-11 guidelines to outperform those of the ICD-10.

Overall, these hypotheses were supported. The more detailed proposed ICD-11 guidelines performed better than the ICD-10 listing, with significant superiority when mental health problems were present. The framing of the individuals’ problems made a difference in whether the RPMs were salient to the rating clinician. Furthermore, correct classification of RPMs was equivalent to that of mental health problems, implying that general mental health clinicians are as adept at recognizing contextual problems as individual ones. It may be that both types of problems are present, clinicians turn to the guidelines to aid in differential diagnosis and that the increased specificity of the proposed ICD-11 guidelines boosts accuracy. WHO’s training materials may want to highlight this inclination to improve accuracy in both multi-problem and family-problem only presentations.

We also found that the application of RPM diagnostic guidelines appears to be related to clinical presentations and not to clinician factors. Classification for the four clinical conditions (i.e., MH presence/absence by RPM presence/absence) was tested for eight demographic and professional factors and none of the 32 tests affected ICD-11 (proposed) or ICD-10 classifications. When ICD-11 and ICD-10 were collapsed, two of the 32 were significant. Differences were related to misclassification of a MH problem rather than to the misclassification of an RPM. Both younger and older clinicians, compared with middle-aged clinicians, were less likely to correctly diagnose the presence of a mental health problem when an RPM was absent. Clinicians from South-East Asia or Africa, compared with clinicians from Europe, were likely to misclassify a MH problem as being absent when only a RPM problem was present. Because this regional effect was for this particular condition, not for RPMs across the board, it could be spurious or it could reveal a regional inclination to

attribute problems to individual, rather than relational, problems where plausible. Although mentioned in the brief training materials clinicians reviewed before rating, WHO may want to underscore this issue, especially with short vignette examples that highlight to clinicians situations in which classification guidelines for both types of problems are met.

Finally, as evidenced both by the patterns of incorrect type of RPM selected and by themes in the clinician feedback on their decisions, clinicians appeared to superimpose a hierarchy on the RPM codes, with relationship distress being less severe than psychological abuse, which is less severe than physical abuse. For example, when behaviors in fact met definitional requirements for psychological abuse but were not perceived as severe enough (as inferred from the open-ended responses justifying their decision-making), clinicians often would assign the relationship distress code, making it a catch-all category for psychological abuse as well as unhappiness.

Should the proposed abuse diagnoses be included in ICD-11, two training implications were brought to the fore by these results. First, the abuse diagnostic guidelines do not require severe harm, a pattern of abuse, an intent to exercise power and control, or any other inference beyond that contained in the field-tested, validated (Heyman, Slep, & Foran, 2015) criteria. Clinicians assess whether a qualifying act (or, in the case of neglect, omission) occurred and whether a qualifying impact was related to or exacerbated by the act/omission. This structure parallels the symptom-with-associated-harm structure of the ICD and DSM (Heyman & Slep, *in press*). Training should highlight both what the guidelines require *and* what they do not. Second and relatedly, many participants appeared to employ heuristics (e.g., representativeness, availability, Tversky & Kahneman, 1974) placing relationship problems on a severity continuum and used these heuristics, rather than the guidelines as written, to make their decisions. Training should approach this problem head-on, alerting clinicians to the error and suggesting heightened attention to the act and impact guidelines that operationalize RPMs.

The modest agreement between clinical diagnoses and the experimenter-fixed “correct” decision in this study for physical and psychological abuse is wholly consistent with similar, non-analogue comparisons in the real world. A meta-analysis (Rettew, Lynch, Achenbach, Dumenci, & Ivanova, 2009) aggregating 38 studies found poor agreement between clinical diagnoses — made with the ICD or with the DSM (American Psychiatric Association, 1980, 1987, 1994) — and those made by independent raters using gold-standard structured clinical interviews. This disparity could be due to clinicians using guidelines differently than intended (as noted in the open-ended responses) or to differences in the decision-making process itself, with clinicians making overall “yes/no” classifications and “gold standard” raters making criterion-by-criterion decisions and then applying the system’s algorithmic rules for overall classification. In the earlier field trials that developed the RPM guidelines (Heyman & Slep, 2006) adapted by the ICD working group, field classifiers, who assessed for and classified only maltreatment, improved from 75% agreement with master reviewers to over 90% agreement when classification switched from an overall decision to criterion-by-criterion

decisions guided by a computerized decision support system, that gathered the classifiers' votes and used the guidelines' logic to make an overall classification. Not only does criterion-by-criterion voting force the rater to attend to each guideline, but also it may change the process from a gestalt decision about whether the case matches implicit archetypes for that type of problem to a feature-based decision about the sub-elements of the clinical presentation. Although criterion-by-criterion decision-making is impractical in many field settings, perhaps it could be incorporated in training to highlight the usefulness of adhering to the written guidelines.

## Limitations

Vignette-based studies provide experimental control of case presentation; however, the case descriptions are necessarily brief and only provide an analogue for ICD classification in real world situations with richer information and access to the informant (Evans et al., 2015). Further, although clinicians could classify cases as meeting or not meeting both mental health and RPM requirements, the universe of mental health and RPM categories presented was limited, narrowing the study's clinical validity. Although the sample of clinicians was international, all possessed sufficient proficiency in English to participate in the study. Therefore, participating clinicians may not be fully representative of clinicians from their respective countries.

## Conclusion

This study used a large, international panel of clinicians to test agreement among clinicians using proposed descriptors for ICD-11 adult Relational Problems and Maltreatment categories. ICD-11 proposed guidelines performed better than those from ICD-10, particularly in the presence of comorbid mental health problems. Correctly identifying the presence of interpersonal violence is particularly important to improving a public health because of significant associated health risks. That ICD-11 guidelines performed better than ICD-10 in their detection in the context of mental disorders is important because it reflects how these issues typically present in clinical practice. This lack of hypothesized superiority may be due to (a) training issues related to some clinicians' decisions being shaded by their archetypes for relationship distress and physical and psychological abuse (e. g., a severity continuum) and to (b) clinical classifications traditionally being made for presence or absence of problems/disorders overall compared with "gold-standard" determinations being made on a criterion-by-criterion basis. These data will be used to revise the RPM training recommendations prior to the release of ICD-11 in 2018.

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Acknowledgments

Thanks to founding members of the Relational Problems and Maltreatment Working Group, Steven Beach, Nadine Kaslow, David Reiss, and Marianne Wamboldt for their yearslong work on the issue and to Martha P. Escobar Lux, Mariangels de Planell Sagner, and Rebeca Robles for the Spanish translation.

## References

- American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Arlington, VA: American Psychiatric Publishing.
- American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd rev.). Arlington, VA: American Psychiatric Publishing.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Arlington, VA: American Psychiatric Publishing.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: American Psychiatric Publishing.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289–300.
- Bray, J. H., & Stanton, M. (Eds.). (2012). *The Wiley-Blackwell Handbook of Family Psychology*. Malden, MA: Wiley-Blackwell.
- Coker, A., Davis, K., Arias, I., Desai, S., Sanderson, M., Brandt, H., & Smith, P. (2002). Physical and mental health effects of intimate partner violence for men and women. *American Journal of Preventive Medicine*, 23, 260–268. [http://dx.doi.org/10.1016/S0749-3797\(02\)00514-7](http://dx.doi.org/10.1016/S0749-3797(02)00514-7)
- Evans, S. C., Roberts, M. C., Keeley, J. W., Blossom, J. B., Amaro, C. M., Garcia, A. M., Stoug, C. O., Canter, K. S., Robles, R., & Reed, G. M. (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. <http://dx.doi.org/10.1016/j.ijchp.2014.12.001>
- Foran, H. M., Beach, S. R. H., Slep, A. M. S., Heyman, R. E., & Wamboldt, M. Z. (Eds.). (2013). *Family Problems and Family Violence: Reliable Assessment and the ICD-11*. New York, NY: Springer.
- Gaut, G., Steyvers, M., Imel, Z. E., Atkins, D. C., & Smyth, P. (2017). Content Coding of Psychotherapy Transcripts Using Labeled Topic Models. *IEEE Journal of Biomedical and Health Informatics*, 21, 476–487. <http://dx.doi.org/10.1109/JBHI.2015.2503985>
- Heyman, R. E., Collins, P. S., Slep, A. M. S., & Knickerbocker, L. (2010). Evidence-based substantiation criteria: Improving the reliability of field decisions of child maltreatment and partner abuse. *Protecting Children*, 25, 35–46.
- Heyman, R. E., & Slep, A. M. S. (2006). Creating and field-testing diagnostic criteria for partner and child maltreatment. *Journal of Family Psychology*, 20, 397–408. <http://dx.doi.org/10.1037/0893-3200.20.3.3>
- Heyman, R. E., & Slep, A. M. S. (2009). Reliability of family maltreatment diagnostic criteria: 41 site dissemination field trial. *Journal of Family Psychology*, 23, 905–910. <http://dx.doi.org/10.1037/a0017011>
- Heyman, R.E., & Slep, A.M. S. (in press). Relational diagnoses and beyond. In Friese B. (Ed.), *APA Handbook of Clinical Psychology*. Washington, DC: American Psychological Association Press.
- Heyman, R. E., Slep, A. M. S., & Foran, H. M. (2015). Enhanced definitions of intimate partner violence for DSM-5 and ICD-11 may

- promote improved screening and treatment. *Family Process*, 54, 64–81. <http://dx.doi.org/10.1111/famp.12121>
- Jacobson, N., Trojanowski, L., & Dewa, C. S. (2012). What do peer support workers do? A job description. *BMC Health Services Research*, 12, 205. <http://dx.doi.org/10.1186/1472-6963-12-205>
- Keeley, J. W., Reed, G. M., Roberts, M. C., Evans, S. C., Medina-Mora, M. E., Robles, R., Rebello, T., Sharan, P., Gureje, O., First, M. B., Andrews, H. F., Ayuso-Mateos, J. L., Gaebel, W., Zielasek, J., & Saxena, S. (2016). Developing a science of clinical utility in diagnostic classification systems: Field study strategies for ICD-11 mental and behavioral disorders. *American Psychologist*, 71, 3–16. <http://dx.doi.org/10.1037/a0039972>
- Keeley, J. W., Reed, G. M., Roberts, M. C., Evans, S. C., Robles, R., Matsumoto, C., Brewin, C. R., Cloitre, M., Perkonig, A., Rousseau, C., Gureje, O., Lovell, A. M., Sharon, P., & Maercker, A. (2016). Disorders specifically associated with stress: A case-controlled field study for ICD-11 mental and behavioural disorders. *International Journal of Clinical & Health Psychology*, 16, 109–127. <http://dx.doi.org/10.1016/j.ijchp.2015.09.002>
- Kiecolt-Glaser, J. K., & Wilson, S. J. (2017). Lovesick: How couples' relationships influence health. *Annual Review of Clinical Psychology*, 13, 421–443. <http://dx.doi.org/10.1146/annurev-clinpsy-032816>
- Lagdon, S., Armour, C., & Stringer, M. (2014). Adult experience of mental health outcomes as a result of intimate partner violence victimisation: A systematic review. *European Journal of Psychotraumatology*, 5, 24794. <http://dx.doi.org/10.3402/ejpt.v5.24794@zept0.2014.5.issue-s3>
- Reed, G. M. (2010). Toward ICD-11: Improving the clinical utility of WHO's International Classification of Mental Disorders. *Professional Psychology: Research and Practice*, 41, 457–464. <http://dx.doi.org/10.1037/a0021701>
- Reed, G. M., Rebello, T. J., Pike, K. M., Medina-Mora, M. E., Gureje, O., Zhao, M., Dai, Y., Roberts, M. C., Maruta, T., Matsumoto, C., Krasnov, V. N., Kulygina, M., Lovell, A. M., Stona, A. C., Sharan, P., Robles, R., Gaebel, W., Zielasek, J., Khoury, B., de Jesus, M., Ayuso-Mateos, J. L., Evans, S. C., Kogan, C. S., & Saxena, S. (2015). WHO's Global Clinical Practice Network for mental health. *The Lancet Psychiatry*, 2, 379–380. [http://dx.doi.org/10.1016/S2215-0366\(15\)00183-2](http://dx.doi.org/10.1016/S2215-0366(15)00183-2)
- Reed, G. M., Roberts, M. C., Keeley, J., Hooppell, C., Matsumoto, C., Sharan, P., Robles, R., Carvalho, H., Wu, C., Gureje, O., & Leal-Leturia, I. (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. <http://dx.doi.org/10.1002/jclp.22031>
- Rettew, D. C., Lynch, A. D., Achenbach, T. M., Dumenci, L., & Ivanova, M. Y. (2009). Meta-analyses of agreement between diagnoses made from clinical evaluations and standardized diagnostic interviews. *International Journal of Methods in Psychiatric Research*, 18, 169–184. <http://dx.doi.org/10.1002/mpr.289>
- Robles, T. F., Slatcher, R. B., Trombello, J. M., & McGinn, M. M. (2014). Marital quality and health: A meta-analytic review. *Psychological Bulletin*, 140, 140–187. <http://dx.doi.org/10.1037/a0031859>
- Schonbrun, Y. C., & Whisman, M. A. (2010). Marital distress and mental health care service utilization. *Journal of Consulting and Clinical Psychology*, 78, 732–736. <http://dx.doi.org/10.1037/a0019711>
- Sullivan, K. T., & Lawrence, E. (Eds.). (2016). *The Oxford Handbook of Relationship Science and Couple Interventions*. New York, NY: Oxford University Press.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1131. <http://dx.doi.org/10.1126/science.185.4157.1124>