RESEARCH ARTICLE

WILEY ADAA
ANXIETY AND DEPRESSION
ASSOCIATION OF AMERICA

Examining the diagnostic utility of the *DSM-5* PTSD symptoms among male and female returning veterans

Jonathan D. Green PhD[1,2] [iD] | Anthony Annunziata PhD[1,2] | Sarah E. Kleiman PhD[1,2] |
Michelle J. Bovin PhD[2,3] | Aaron M. Harwell BA[1] | Annie M.L. Fox BA[1*] |
Shimrit K. Black PhD[1,2] | Paula P. Schnurr PhD[4,5] | Darren W. Holowka PhD[6,7] |
Raymond C. Rosen PhD[8] | Terence M. Keane PhD[2,3] | Brian P. Marx PhD[2,3]

[1]VA Boston Healthcare System, Boston, MA, USA

[2]Department of Psychiatry, Boston University School of Medicine, Boston, MA, USA

[3]National Center for PTSD at the VA Boston Healthcare System, Boston, MA, USA

[4]Executive Division, National Center for PTSD, White River Junction, VT, USA

[5]Department of Psychiatry, Geisel School of Medicine at Dartmouth, Hanover, NH, USA

[6]Department of Psychiatry, Rhode Island Hospital, Providence, RI, USA

[7]Department of Psychiatry and Human Behavior, Alpert Medical School of Brown University

[8]New England Research Institutes, Watertown, MA, USA

Correspondence
Brian P. Marx, VA Boston Healthcare System, National Center for PTSD - Behavioral Science Division, 150 South Huntington Avenue (116B-4), Boston, MA 02130.
Email: Brian.Marx@VA.gov

*Annie M. Ledoux is currently at George Mason University.

Background: Posttraumatic stress disorder (PTSD) diagnostic criteria have been criticized for including symptoms that overlap with commonly comorbid disorders, which critics argue undermines the validity of the diagnosis and inflates psychiatric comorbidity rates. In response, the upcoming 11th edition of the International Classification of Diseases (ICD-11) will offer PTSD diagnostic criteria that are intended to promote diagnostic accuracy. However, diagnostic utility analyses have not yet assessed whether these criteria minimize diagnostic errors. The present study examined the diagnostic utility of each PTSD symptom in the fifth edition of the *Diagnostic and Statistical Manual for Mental Disorders* (*DSM-5*) for males and females.

Methods: Participants were 1,347 individuals enrolled in a longitudinal national registry of returning veterans receiving care at a Department of Veterans Affairs (VA) facility. Doctoral level clinicians assessed all participants using the PTSD module of the Structured Clinical Interview for *DSM*.

Results: Of the 20 symptoms examined, the majority performed in the fair to poor range on test quality indices. Although a few items did perform in the good (or better) range, only half were ICD-11 symptoms. None of the 20 symptoms demonstrated good quality of efficiency. Results demonstrated few sex differences across indices. There were no differences in the proportion of comorbid psychiatric disorders or functional impairment between *DSM-5* and ICD-11 criteria.

Conclusions: ICD-11 PTSD criteria demonstrate neither greater diagnostic specificity nor reduced rates of comorbidity relative to *DSM-5* criteria and, as such, do not perform as intended. Modifications to existing symptoms or new symptoms may improve differential diagnosis.

KEYWORDS
Diagnostic and Statistical Manual for Mental Disorders, diagnostic techniques and procedures, International Classification of Diseases, posttraumatic, psychological trauma, stress disorder

1 | INTRODUCTION

Since their introduction, the diagnostic criteria for posttraumatic stress disorder (PTSD) in the fifth edition of the *Diagnostic and Statistical Manual for Mental Disorders* (*DSM-5*; American Psychiatric Association [APA], 2013) have been criticized. One concern is that the new symptoms, including negative beliefs about oneself or the world (symptom D2), strong negative emotions (D4), irritable or aggressive behavior (E1), and reckless or self-destructive behavior (E2), may overlap with symptoms of commonly comorbid disorders (Brewin, Lanius,

Novac, Schnyder, & Galea, 2009; Hoge et al., 2016; Rosen & Lilienfeld, 2008; Rosen, Spitzer, & McHugh, 2008). Critics contend that including syndromally indistinct symptoms like these may undermine the validity of the diagnosis and inflate comorbidity rates (Rosen, Lilienfeld, Frueh, McHugh, & Spitzer, 2010; Spitzer, Rosen, & Lilienfeld, 2008).

To address this concern, some have suggested eliminating overlapping symptoms from the diagnosis (Spitzer, First, & Wakefield, 2007). Consistent with this perspective, the proposed PTSD criteria for the 11th edition of the International Classification of Diseases (ICD-11) includes only six symptoms (see Fig. 1), chosen based on

PTSD Symptom Clusters for DSM-III-R, DSM-IV, DSM-IV-TR, and DSM-5

DSM-5	DSM-IV-TR	DSM-III-R
Criterion A: Stressor	Criterion A: Stressor	Criterion A: Stressor
A1. Exposure to actual or threatened death, serious injury, or sexual violence in one (or more) of the following ways: 1. Directly experiencing the traumatic event(s). 2. Witnessing, in person, the event(s) as it occurred to others. 3. Learning that the traumatic event(s) occurred to a close family member or close friend. *Event(s) must have been violent or accidental.* 4. Experiencing repeated or extreme exposure to aversive details of the traumatic event(s). *Media does not count.* A2. Eliminated	A1. The person experienced, witnessed, or was confronted with an event that involved actual or threatened death or serious injury or a threat to the physical integrity of self or others. A2. The response involved intense fear, helplessness, or horror.	A1. The person has experienced an event outside the range of usual human experience and that would be markedly distressing to almost anyone.
Cluster B (Intrusion)	Cluster B (Intrusive Recollection)	Cluster B (Intrusion)
B1. Intrusive memories	B1. Intrusive recollections	B1. Intrusive recollections
B2. Nightmares[a]	B2. Nightmares	B2. Nightmares
B3. Flashbacks[a]	B3. Flashbacks	B3. Flashbacks
B4. Psychological distress at exposure to cues	B4. Psychological distress at exposure to cues	B4. Psychological distress at exposure to cues
B5. Physiological reactivity at exposure to cues	B5. Physiological reactivity at exposure to cues	
Cluster C (Avoidance)	Cluster C (Avoidance and Numbing)	Cluster C (Avoidance and Numbing)
C1. Avoidance of thoughts and feelings[a]	C1. Avoid thoughts and feelings	C1. Avoid thoughts and feelings
C2. Avoidance of external reminders[a]	C2. Avoid reminders	C2. Avoid reminders
Cluster D (Negative changes in cognitions and mood)		
D1. Inability to recall important aspect of trauma	C3. Psychogenic amnesia	C3. Psychogenic amnesia
D2. Negative beliefs about oneself or the world[b]	C7. Foreshortened future	C7. Foreshortened future
D3. Distorted self-blame or other-blame		
D4. Persistent negative emotions[b]		
D5. Diminished interest in significant activities[b]	C4. Anhedonia	C4. Anhedonia
D6. Feelings of detachment or estrangement	C5. Detached from others	C5. Detached from others
D7. Inability to experience positive emotions	C6. Psychic numbing	C6. Psychic numbing
Cluster E (Changes in arousal and reactivity)	Cluster D (Hyper-arousal)	Cluster D (Hyper-arousal)
E1. Irritable or aggressive behavior[b]	D2. Irritability/anger	D2. irritability or outbursts of anger
E2. Reckless or self-destructive or behavior[b]		
E3. Hypervigilance[a]	D4. Hypervigilance	D4. hypervigilance
E4. Exaggerated startle response[a]	D5. Exaggerated startle	D5. exaggerated startle response
E5. Problems with concentration	D3. Concentration difficulty	D3. difficulty concentrating
E6. Sleep disturbance	D1. Disturbed sleep	D1. difficulty falling or staying asleep
		D6. physiologic reactivity
Dissociative Subtype		
Derealization		
Depersonalization		

Note: DSM-III-R and DSM-IV-R are aligned horizontally with DSM-5 counterparts.
[a]Symptom is present in both DSM-5 and ICD-11.
[b]Symptom commonly overlaps with comorbid disorders.

FIGURE 1 PTSD Symptom Clusters for DSM-III-R, DSM-IV, DSM-IV-TR, and DSM-5

the assumption that they are unique to PTSD (Cloitre, Garvert, Brewin, Bryant, & Maercker, 2013; Maercker et al., 2013). The ICD-11 committee reasoned that a diagnosis including only "core symptoms" would improve upon *DSM-5* criteria by enhancing diagnostic accuracy, reducing overlap with comorbid psychiatric conditions, and decreasing assessment time and burden (Brewin, 2013; Brewin et al., 2009; Maercker et al., 2013).

Although these justifications are compelling, determining if symptoms chosen to represent PTSD in ICD-11 are the most specific requires an examination of the diagnostic utility of these and other *DSM-5* PTSD symptoms. Diagnostic utility analyses provide information about the sensitivity (probability of endorsement among patients with the diagnosis), specificity (probability of lack of endorsement among patients without the diagnosis), and efficiency (probability that endorsement corresponds to diagnostic status) of each symptom (Kraemer, 1992). The ICD-11 approach of including only unique symptoms emphasizes diagnostic specificity.

In the only study to examine the diagnostic utility of each PTSD symptom, Holowka, Marx, Kaloupek, and Keane (2012) tested *DSM-III-R* (APA, 1987) PTSD diagnostic criteria in a large sample of male Vietnam veterans, finding that both unique and overlapping symptoms had high levels of sensitivity (e.g., hypervigilance and startle) and specificity (e.g., nightmares and difficulty concentrating). Although the authors

noted that intrusive memories and detachment from others most accurately predicted the overall PTSD diagnosis, diagnostic efficiency was not reported.

In this study, we examined the diagnostic utility of each *DSM-5* PTSD symptom, expanding on Holowka et al. (2012) work in several ways. First, we included both male and female participants because research suggests that men and women may have different PTSD symptom profiles (Fullerton et al., 2001; Green, 2003; Zlotnick, Zimmerman, & Wolfsdorf, 2001), suggesting that the diagnostic utility of each PTSD symptom may vary by gender. Second, as measures of test performance (i.e., sensitivity, specificity, efficiency) could be inflated due to the high prevalence of PTSD in our sample, we examined measures of test quality (quality of sensitivity, QSN; quality of specificity, QSP; quality of efficiency, QEF), which are superior to measures of test performance because they calibrate for chance agreement between test and diagnosis (Kraemer, 1992). Third, we explored the efficiency of each PTSD item. Finally, we examined whether ICD-11 criteria reduced psychiatric comorbidity.

Consistent with Holowka et al. (2012), of the six ICD-11 PTSD symptoms, we hypothesized that nightmares (B2) and hypervigilance (E3) would demonstrate the highest QSN, and that nightmares (B2) and flashbacks (B3) would demonstrate the highest QSP. Based on the ICD-11 committee's rationale, we hypothesized that ICD-11

symptoms would demonstrate strong QSP. Similarly, we expected a lower proportion of comorbid disorders among participants who met criteria for PTSD under ICD-11 criteria than those that met under *DSM-5*. Consistent with Holowka et al., we hypothesized that symptoms both unique to PTSD and overlapping with other disorders would demonstrate diagnostic utility. However, because PTSD diagnostic criteria have changed from *DSM-III-R* to *DSM-5*, and because our analytic strategy differed from Holowka et al., we did not have any a priori hypotheses regarding the diagnostic utility of these additional symptoms. Finally, because no study has examined the diagnostic utility of each PTSD symptom by gender, our examination of gender differences was exploratory.

2 | MATERIALS AND METHODS

2.1 | Participants

Participants were veterans enrolled in the Veterans After-Discharge Longitudinal Registry (Project VALOR), a longitudinal national registry of Operation Enduring Freedom and Operation Iraqi Freedom (OEF/OIF) Army and Marine Corps veterans (Rosen et al., 2012). To be included in Project VALOR, veterans must have undergone a mental health evaluation at a VA facility. Veterans with probable PTSD according to VA medical records (at least two instances of a PTSD diagnosis by a mental health professional associated with two separate visits) were oversampled to create a 3:1 (PTSD:no PTSD) ratio. Veterans without any PTSD diagnoses during the same time frame were eligible to be included in the no PTSD group. Veterans with just one PTSD diagnosis during the same window were excluded. As the registry was assembled between July 2008 and December 2009, diagnoses were made using *DSM-IV* (APA, 1994) criteria. Females were oversampled to create a 1:1 sex ratio. Potential Project VALOR participants were recruited from a roster of veterans, provided by the VA Environmental Epidemiology Service, who met inclusion criteria. Potential participants ($n = 4,331$) were contacted by phone. Of these, 2,712 (62.6%) consented to participate. Of those, 2,169 (80.0%) completed study questionnaires and 1,649 (60.8%) completed both the questionnaires and the diagnostic interview. At that time (Time 1 [T1], December 2009 to September 2012), *DSM-IV* criteria were available and the Structured Clinical Interview for *DSM-IV* (SCID-IV; First, Spitzer, Gibbon, & Williams, 1996) was used during this interview. Data from these 1,649 veterans were included in Project VALOR as T1 data. At T1, 75.9% of participants met criteria for PTSD on the SCID-IV. The present study consisted of 1,347 veterans who participated in Time 2 (T2; September 2013 to August 2014) of Project VALOR. At that time, *DSM-5* criteria were available and the SCID-5 (First, Williams, Karg, & Spitzer, 2015) was used to assess these participants.

2.2 | Measures

2.2.1 | Demographics

Participants completed a demographics questionnaire online or by mail. They reported their age, sex, and race.

2.2.2 | PTSD diagnosis

The SCID-5 PTSD module was used to assess current (past month) PTSD symptoms and diagnostic status. Though the SCID-5 for PTSD has not been evaluated for psychometric properties, *DSM-5* PTSD criteria demonstrated good reliability ($\kappa = .69$) in field trials (Regier et al., 2013). Of note, PTSD prevalence was higher in the *DSM-5* field trials than in most populations, potentially resulting in an inflated kappa. In the present study, interrater agreement was excellent ($\kappa = .82$) among a random subset of 100 interviews that were rated by an assessor who did not complete the initial interview.

2.2.3 | Depression diagnosis

The SCID-5 MDD module was used to assess for current (past month) MDD diagnostic status. The SCID-5 for MDD has not been evaluated for psychometric properties, and *DSM-5* criteria has demonstrated questionable reliability ($\kappa = .20–.35$; Regier et al., 2013). In the present study, assessment of interrater agreement for the MDD module was identical to that for the PTSD module, and was excellent ($\kappa = .75$).

2.2.4 | Alcohol use disorder

The Alcohol Use Disorders Identification Test (AUDIT) is a 10-item questionnaire that was used to classify participants with problematic alcohol use (Saunders, Aasland, Babor, De la Fuente, & Grant, 1993). In this study, Cronbach's α was .87. A cut-score of 8 was used to indicate hazardous and harmful alcohol use (Babor, Higgins-Biddle, Saunders, & Monteiro, 2001; Conigrave, Hall, & Saunders, 1995).

2.2.5 | Panic syndrome and generalized anxiety disorder

The PHQ is a self-report version of the Primary Care Evaluation of Mental Disorders (PRIME-MD; Spitzer, Kroenke, & Williams, 1999). It is a 58-item measure that can be used to determine probable diagnostic status for several mental disorders. We used scoring procedures outlined in the measure's manual to create dichotomous diagnostic categories for panic syndrome and generalized anxiety disorder (GAD). Cronbach's α was .72 and .85 for panic syndrome and GAD, respectively.

2.2.6 | Functional impairment

The inventory of psychosocial functioning (IPF) is an 80-item self-report measure assessing PTSD-related functional impairment (Rodriguez, Holowka, & Marx, 2012). It yields an overall score of psychosocial impairment, with higher scores indicating greater impairment. In this study, Cronbach's α was .73. We used a cutoff of 51 to indicate psychosocial impairment, which is indicative of "severe impairment" (Bovin et al., in press).

2.3 | Procedure

Participants completed questionnaires online or by mail and were then interviewed via telephone by doctoral level clinicians. All participants completed T1 of Project VALOR ~2.5 years prior to the current assessment. Participants provided informed consent prior to participation. The study was approved by the local Institutional Review Boards and the Human Research Protection Office of the US Army

Medical Research and Materiel Command. Participants were compensated $100 for completing T2 of Project VALOR.

2.4 | Analyses

Using SPSS version 24, we computed percentages of participants who met criteria for each PTSD symptom based on the SCID-5, as well as the number and percent of participants who met criteria for PTSD based on both *DSM-5* and ICD-11 criteria (IBM Corp, 2016). We compared the prevalence of these diagnoses via comparative error (CE) analyses. CE was calculated using the following equation:

$$CE = 1.96x\sqrt{(r1(100 - r1) \div s1) + (r2(100 - r2) \div s2)}.$$

In this equation, $r1$ is the percentage response for the first group, $r2$ is percentage response for the second group, $s1$ is the sample size of the first group, and $s2$ is the sample size of the second group. Significance was calculated using an online calculator (EasyCalculation.com).

For diagnostic utility analyses, we created 2×2 contingency tables to classify participants based on the presence/absence of each PTSD symptom and presence/absence of PTSD diagnosis. We analyzed one 2×2 table for each of the 20 PTSD symptoms. Each table classified participants into one of four cells: true positives (symptom+ and diagnosis+), true negatives (symptom− and diagnosis−), false positives (symptom+ and diagnosis−), and false negatives (symptom− and diagnosis+). To avoid conditional dependence issues, the 2×2 table for each symptom was created based on a PTSD diagnosis that was calculated without that symptom (e.g., the B1 2×2 table was created using a PTSD diagnosis that was calculated with B1 excluded). For each 2×2 table, three measures of test performance (sensitivity, specificity, and efficiency) and three measures of test quality were calculated using DAG_STAT software (Mackinnon, 2000). Test quality measures were weighted κ coefficients as proposed by Kraemer (1992) for QSN ($\kappa[1]$), QSP ($\kappa[0]$), and QEF ($\kappa[.5]$). We judged the clinical significance of κ coefficients using Cicchetti (1994)'s guidelines: $\kappa \leq .40$ is poor, $\kappa \geq .41$ and $< .60$ is fair, $\kappa \geq .60$ and $< .75$ is good, and $\kappa \geq .75$ is excellent. We conducted all analyses separately for males and females.

To determine whether ICD-11 criteria reduce comorbidity, we calculated the CE between the proportion of those with a *DSM-5* PTSD diagnosis and a comorbid disorder (e.g., alcohol use disorder [AUD]) and those with an ICD-11 PTSD diagnosis and the same comorbidity.

3 | RESULTS

Participant demographics are reported in Table 1. PTSD diagnostic prevalence was not significantly different between ICD-11 and *DSM-5* criteria, with 846 (62.8%) participants meeting criteria for *DSM-5* PTSD and 874 (64.9%) for ICD-11 PTSD (CE = 3.63, *ns*). Fifty-one (3.8%) participants met criteria for *DSM-5* but not ICD-11, and 79 (5.9%) met criteria for ICD-11 but not *DSM-5*. One hundred thirty (9.6%) cases were discrepant between ICD-11 and *DSM-5*. There were no differences in the proportion of comorbid depression (CE = 4.48, *n.s.*), GAD (CE = 4.54, *n.s.*), panic syndrome (CE = 4.70, *n.s.*), AUD (CE

= 4.17, *n.s.*), or functional impairment (CE = 1.68, *n.s.*) between those meeting criteria under *DSM-5* versus ICD-11.

Regarding quality indices, items demonstrated similar patterns across gender. Only four items achieved good QSN. Three of these are considered "core" PTSD symptoms (intrusive memories [B1], avoidance of external reminders [C2], and hypervigilance [E3]); the fourth is not (feelings of detachment or estrangement [D6]). Of note, intrusive memories (B1) demonstrated good QSN for men, but only fair QSN for women.

Two symptoms had, at minimum, good QSP for both men and women; inability to experience positive emotions (D7) demonstrated good QSP for both genders, while flashbacks (B3) demonstrated excellent QSP for women and good QSP for men. Distorted self- or other-blame (D3) and persistent negative emotions (D4) had good QSP for women and fair QSP for men. Reckless or self-destructive behavior (E2) demonstrated good QSP among men and fair QSP among women. Only one ICD-11 symptom demonstrated good QSP (flashbacks [B3]).

None of the 20 symptoms assessed had good or excellent QEF. Eight symptoms demonstrated fair QEF for both genders. Three additional symptoms had fair QEF for men but poor QEF for women: persistent negative emotions (D4), exaggerated startle response (E4), and sleep disturbance (E6). In contrast, avoidance of thoughts and feelings (C1) demonstrated fair QEF for women but poor QEF for men (see Table 2).

Two symptoms did not perform well across test quality indices. Both inability to recall important aspects of the trauma (D1) and irritable or aggressive behavior (E1) demonstrated poor diagnostic utility across all three indices.[1]

4 | DISCUSSION

Contrary to hypotheses, the only proposed ICD-11 symptom that demonstrated good (for males) or excellent (for females) QSP was flashbacks (B3). All other ICD-11 symptoms demonstrated fair or poor QSP. Also contrary to hypotheses, there were no differences in rates of comorbid psychiatric disorders or psychosocial functioning between those who met criteria for PTSD under *DSM-5* versus ICD-11. These findings suggest that the proposed ICD-11 PTSD criteria may not perform as anticipated.

Hypotheses stemming from work done by Holowka et al. (2012) were partially supported. Both symptoms unique to PTSD and those thought to overlap with comorbid disorders demonstrated good diagnostic utility. Although our analyses do provide guidance as to the most specific PTSD symptoms, creating a diagnosis using only these items is not recommended. A definition including only specific items, even those with excellent QSP, would inflate the number of false negatives. Ideally, a definition would include a mix of items with high QSP, QSN, and QEF. Inclusion of diagnostically efficient symptoms is particularly important because they minimize diagnostic errors. Therefore, diagnostically efficient, rather than specific, symptoms would be most likely to separate PTSD from other commonly comorbid disorders.

In this study, none of the *DSM-5* or ICD-11 PTSD symptoms demonstrated good or excellent QEF. This does not necessarily suggest that the *DSM-5* and ICD-11 diagnostic conceptualizations are inherently

TABLE 1 Demographic information

	Females (n = 689)	Males (n = 658)	All Participants (n = 1347)
Age—M (SD)	40.01 (9.36)	41.40 (10.14)	40.69 (9.77)
Race/ethnicity (%)			
White	487 (70.7)	528 (80.2)	75.4
Black	147 (21.3	71 (10.8)	16.8
American Indian or Alaska Native	18 (2.6)	20 (3.0)	2.8
Asian	18 (2.6)	6 (.9)	1.8
Pacific Islander	4 (.6)	6 (.9)	0.7
Other Race	15 (2.2)	27 (4.1)	3.1
Hispanic	89 (12.9)	79 (12.0)	12.4

flawed. That a number of symptoms demonstrated good to excellent QSN and QSP indicates that *DSM-5* and ICD-11 PTSD diagnostic criteria include symptoms that, when used in combination, may adequately detect the presence or absence of PTSD. Unfortunately, results of this study cannot provide guidance on which symptom combination is optimal. It is possible that delineating symptoms that achieve good to excellent QEF may be accomplished by modifying wording of existing criteria, especially for symptoms that exhibited the highest QEF.

It is also possible that, despite modification, no current PTSD symptoms can yield good or excellent QEF. Instead, there may be symptoms that are not part of any classification system that better distinguish PTSD from other disorders. PTSD assessment instruments such as the Mississippi Scale for Combat-Related PTSD (Keane, Caddell, & Taylor, 1988) and the Detailed Assessment of Posttraumatic Stress (Briere, 2001) include more items than just those that directly correspond to *DSM* diagnostic criteria. The items included on these and other scales might be worth investigating as diagnostic indicators.

Rather than an issue of incomplete content, our inability to identify items demonstrating good or excellent QEF could be due to measurement error. The SCID-5 has not been validated and does not encourage the same in-depth probing as other PTSD interviews (i.e., the Clinician Administered PTSD Scale for *DSM-5*; Weathers et al., 2013). Symptoms with only fair QEF in the current study might demonstrate good QEF with a different diagnostic tool. This possibility awaits empirical exploration.

Although many of our findings were consistent with Holowka et al. (2012) investigation of *DSM-III-R* PTSD symptoms, there were notable differences. However, Holowka et al. found that nightmares and physiological reactivity to cues were diagnostically specific, these symptoms demonstrated only fair QSP in our study. Such differences may reflect the fact that Holowka et al. examined test performance rather than test quality in a sample comprised entirely of male Vietnam War veterans, rather than of both sexes and who served in more recent conflicts. The broad similarities found across the two studies, however, suggest that despite significant revisions to the PTSD criteria between *DSM-III-R* and *DSM-5*, symptom performance is consistent.

Findings were generally consistent across genders. The five symptoms with the highest kappas across all three quality indices tended to be the same for men and women. Although males and females may differ somewhat in PTSD symptom presentation, the same symptoms seem to signal the presence and absence of PTSD in both sexes.

Our findings have important clinical implications beyond diagnostic classification. The identification of symptoms with high QSN, which minimize false negatives, has implications for screening tools. One common PTSD screening tool, the primary care PTSD screen (PC-PTSD; Prins et al., 2004), was recently revised for *DSM-5* (Prins et al., 2016). Our results suggest that the PC-PTSD-5 includes items with the highest QSN (e.g., avoidance of external reminders [C2] and hypervigilance [E3]). However, it also includes distorted self- or other-blame (D3), which has good QSP but poor QSN. Application of our findings to the PC-PTSD-5 may be limited as the PC-PTSD-5 was designed for use in primary care settings. However, this example highlights the relevance of our study to screening tool evaluations.

Similarly, the identification of symptoms with high QSP (e.g., flashbacks [B3]), distorted self- or other-blame [D3], and persistent negative emotions [D4]), which decrease false positives, could be useful for clinicians with limited resources hoping to confirm a PTSD diagnosis (Kraemer, 1992). Future work is needed to examine whether truncated confirmatory assessment tools could be developed.

Both irritable or aggressive behaviors (E1) and inability to recall important aspects of the trauma (D1) demonstrated poor diagnostic utility across all indices. Findings for D1 are consistent with the broader literature (Armour et al., 2015; Holowka et al., 2012; Keane et al., 2014). Therefore, it may be appropriate to remove these symptoms from the diagnosis.

Study findings should be viewed in light of limitations. First, the high prevalence of PTSD in this sample likely resulted in a deflated estimation of false positives, which could underrepresent the number of discrepant cases between *DSM-5* and ICD-11 diagnoses. In more representative samples of OEF/OIF veterans, in which PTSD prevalence ranges from 15 to 20% (Ramchand et al., 2010), the corresponding increase in false positives would result in a higher proportion of discrepant cases. This is a limitation of other recent work comparing ICD-11 and *DSM-5* criteria as well (e.g., Hafstad, Thoresen, Wentzel-Larsen, Maercker, & Dyb, 2017). Future research should investigate these differences in samples with PTSD prevalence rates that are comparable to populations of interest. Second, in the present study, both *DSM-5* and ICD-11 were assessed using the SCID-5. Ideally, these criteria would

TABLE 2 Endorsement of PTSD symptoms and diagnostic utility analyses for DSM-5

Symptom	PTSD Status Absent %Men/%Women ($n_{men}=244$)($n_{women}=257$)	Present %Men/%Women ($n_{men}=414$)($n_{women}=432$)	Sens Men/Women	Spec Men/Women	Eff Men/Women	κ(0)[a] Men/Women	κ(.5)[b] Men/Women	κ(1)[c] Men/Women
Cluster B (intrusion)								
B1. Intrusive memories[d]	50.0/40.2	92.5/89.1	.92/.88	.47/.56	.76/.77	.32/.39	.43/.47	.65/.57
B2. Nightmares[e]	28.2/23.2	71.5/73.6	.72/.72	.73/.75	.72/.73	.53/.56	.42/.45	.35/.37
B3. Flashbacks[e]	5.9/4.4	25.6/28.7	.25/.28	.94/.96	.48/.50	.70/.79	.14/.17	.08/.10
B4. Psychological distress at exposure to cues[d]	31.4/29.4	76.8/76.9	.77/.77	.69/.73	.74/.76	.50/.56	.45/.48	.40/.42
B5. Physiological reactivity at exposure to cues[d]	28.3/30.1	77.8/80.3	.77/.79	.71/.70	.75/.76	.53/.52	.46/.48	.41/.45
Cluster C (avoidance)								
C1. Avoidance of thoughts and feelings[e]	40.5/41.0	84.3/88.4	.83/.88	.54/.56	.72/.76	.33/.38	.38/.46	.45/.56
C2. Avoidance of external reminders[e]	43.6/52.8	93.0/94.9	.92/.94	.45/.40	.71/.72	.27/.25	.39/.37	.67/.70
Cluster D (negative changes in cognitions and mood)								
D1. Inability to recall important aspect of the trauma[d]	13.6/15.7	23.2/33.8	.22/.31	.83/.80	.43/.48	.15/.27	.04/.09	.02/.05
D2. Negative beliefs about oneself or the world[d]	30.6/34.2	79.0/81.9	.78/.80	.66/.61	.74/.74	.45/.41	.43/.42	.42/.43
D3. Distorted self- or other-blame	10.5/18.5	52.7/61.3	.51/.60	.87/.80	.63/.67	.66/.56	.32/.35	.21/.25
D4. Persistent negative emotions[d]	14.1/17.5	64.0/66.2	.63/.64	.84/.79	.70/.69	.66/.58	.41/.39	.30/.29
D5. Diminished interest in significant activities[d]	20.9/23.5	78.7/76.4	.78/.79	.75/.72	.77/.74	.58/.53	.51/.46	.46/.40
D6. Feelings of detachment or estrangement	33.6/36.1	89.9/90.0	.89/.89	.62/.58	.79/.78	.46/.41	.53/.49	.62/.61
D7. Inability to experience positive emotions[d]	14.5/14.0	55.3/49.8	.55/.49	.85/.86	.65/.61	.63/.61	.33/.28	.23/.18
Cluster E (changes in arousal and reactivity)								
E1. Irritable or aggressive behavior[d]	37.4/38.9	68.8/70.8	.69/.70	.65/.61	.68/.67	.40/.34	.32/.29	.27/.25
E2. Reckless or self-destructive behavior[d]	9.5/5.2	29.2/22.7	.29/.22	.91/.94	.49/.46	.59/.67	.15/.12	.08/.06
E3. Hypervigilance[e]	72.7/63.8	95.4/95.4	.95/.95	.28/.34	.72/.73	.17/.22	.27/.33	.63/.65
E4. Exaggerated startle response[e]	40.0/52.0	83.8/86.3	.83/.85	.61/.46	.76/.72	.43/.27	.45/.33	.47/.42
E5. Problems with concentration[d]	40.5/39.3	84.5/84.3	.84/.83	.59/.58	.76/.74	.42/.40	.44/.42	.47/.44
E6. Sleep disturbance[d]	43.4/50.4	84.5/87.2	.85/.86	.58/.47	.76/.72	.40/.29	.44/.35	.48/.45

PTSD, posttraumatic stress disorder; Sens, sensitivity; Spec, specificity; Eff, efficiency.

[a] κ(0) = quality of specificity.
[b] κ(.5) = quality of efficiency.
[c] κ(1) = quality of sensitivity.
[d] Symptom commonly overlaps with those in comorbid disorders.
[e] Symptom is present in both DSM-5 and ICD-11.

TABLE 3 Endorsement of PTSD symptoms and diagnostic utility analyses for ICD-11

Symptom	PTSD status Absent %Men/%Women ($n_{men}=291$) ($n_{women}=299$)	Present %Men/%Women ($n_{men}=344$) ($n_{women}=361$)	Sensitivity Men/Women	Specificity Men/Women	Efficiency Men/Women	$\kappa(0)$[b] Men/Women	$\kappa(.5)$[c] Men/Women	$\kappa(1)$[d] Men/Women
Cluster B (intrusion)								
B1. Intrusive memories[e]	57.7/50.8	94.8/89.6	.95/.90	.42/.49	.71/.71	.26/.30	.39/.40	.76/.63
B2. Nightmares[a]	10.0/8.4	95.9/95.6	.88/.88	.50/.52	.57/.59	.12/.14	.21/.23	.72/.72
B3. Flashbacks[a]	1.7/1.0	33.4/36.3	.31/.33	.94/.94	.61/.62	.67/.72	.24/.27	.14/.17
B4. Psychological distress at exposure to cues[e]	40.5/38.5	78.5/78.4	.78/.79	.59/.61	.70/.71	.34/.36	.38/.41	.45/.46
B5. Physiological reactivity at exposure to cues[e]	37.6/39.1	80.2/82.5	.80/.83	.62/.61	.72/.73	.38/.38	.43/.44	.50/.53
Cluster C (avoidance)								
C1. Avoidance of thoughts and feelings[a]	46.4/51.5	88.4/88.9	.87/.88	.49/.46	.69/.68	.26/.25	.37/.35	.59/.59
C2. Avoidance of external reminders[a]	56.0/62.2	92.7/95.3	.92/.95	.39/.33	.64/.63	.19/.17	.30/.28	.66/.73
Cluster D (negative changes in cognitions and mood)								
D1. Inability to recall important aspect of the trauma	15.5/22.7	23.8/31.3	.24/.31	.84/.77	.52/.52	.23/.17	.08/.08	.05/.05
D2. Negative beliefs about oneself or the world[e]	46.4/51.2	75.6/77.0	.66/.77	.65/.49	.66/.64	.36/.22	.30/.27	.25/.34
D3. Distorted self- or other-blame	22.7/30.8	51.2/59.3	.51/.59	.77/.69	.63/.64	.41/.33	.28/.28	.21/.24
D4. Persistent negative emotions[e]	29.2/33.1	61.6/62.6	.62/.63	.71/.67	.66/.65	.38/.33	.32/.30	.28/.27
D5. Diminished interest in significant activities[e]	41.9/43.5	73.0/69.8	.73/.70	.58/.56	.66/.64	.28/.25	.31/.26	.34/.28
D6. Feelings of detachment or estrangement[e]	56.7/56.0	82.0/83.9	.82/.84	.43/.44	.64/.66	.19/.21	.26/.29	.39/.44
D7. Inability to experience positive emotions[e]	28.5/26.4	52.0/46.3	.52/.46	.71/.74	.61/.59	.31/.29	.23/.19	.19/.14
Cluster E (changes in arousal and reactivity)								
E1. Irritable or aggressive behavior[e]	44.0/49.2	69.5/68.7	.69/.69	.56/.51	.63/.61	.24/.18	.26/.20	.28/.22
E2. Reckless or self-destructive behavior[e]	13.7/13.7	29.7/19.1	.30/.19	.86/.86	.56/.50	.39/.17	.15/.05	.09/.03
E3. Hypervigilance[a]	75.3/69.9	98.0/96.4	.98/.96	.21/.27	.57/.60	.10/.13	.18/.23	.81/.74
E4. Exaggerated startle response[a]	47.8/56.9	86.3/88.9	.86/.89	.51/.41	.70/.66	.29/.21	.38/.31	.56/.55
E5. Problems with concentration[e]	55.0/54.8	81.4/80.0	.82/.80	.45/.45	.65/.64	.21/.20	.27/.26	.40/.36
E6. Sleep disturbance[e]	53.4/58.5	84.6/87.3	.85/.88	.47/.41	.67/.67	.24/.21	.33/.30	.50/.51

PTSD, posttraumatic stress disorder; Sens, sensitivity; Spec, specificity; Eff, efficiency.

[a]Symptom is present in both DSM-5 and ICD-11.

[b]$\kappa(0)$ = quality of specificity.

[c]$\kappa(.5)$ = quality of efficiency.

[d]$\kappa(1)$ = quality of sensitivity.

[e]Symptom commonly overlaps with those in comorbid disorders.

have been assessed independently. However, there currently exists no standardized assessment for ICD-11 PTSD.

5 | CONCLUSION

Our results suggest that the *DSM-5* PTSD criteria include symptoms with strong QSN and QSP. This mix is important for limiting false positives and false negatives(Kraemer, 1992). However, none of the current symptoms demonstrated strong QEF. Contrary to hypotheses, our findings suggest that ICD-11 criteria may not perform as intended; items chosen for presumed QSP performed otherwise. Further, participants diagnosed with PTSD under each set of diagnostic criteria exhibited similar rates of psychiatric comorbidities and similar levels of functional impairment. As ICD-11 offers a briefer set of symptoms, it may be more convenient to use ICD-11 criteria in situations where diagnoses must be made quickly. In contrast, assessments using *DSM-5* criteria may offer information regarding treatment targets (e.g., distorted self- or other-blame [D4] for use in cognitive processing therapy; Resick, Monson, & Chard, 2006). However, only 10% of cases were discrepant between *DSM-5* and ICD-11 in this study, this number will increase as prevalence decreases. Therefore, choice of diagnostic classification system has important implications for over- and underdiagnosis. We encourage future research to use these findings as a starting point for garnering a better understanding of the PTSD construct.

ENDNOTE

[1] We conducted the same analyses using the ICD-11 PTSD diagnosis as the comparison. Three diagnostic utility scores remained the same as when compared to the *DSM-5* PTSD diagnosis, 12 scores improved but not enough to affect their level of clinical significance, and two scores (QSN for men on psychological distress at exposure to cues [B2] and hypervigilance [E3]) increased in level of clinical significance, from poor to fair and poor to excellent, respectively. Of these, only hypervigilance is included in the ICD-11 diagnostic criteria. Diagnostic utility was consistently poorer when compared to the ICD-11 diagnosis (see Table 3).

REFERENCES

American Psychiatric Association (APA). (1987). *Diagnostic and statistical manual of mental disorders, revised third edition (DSM-III)*. Washington, DC: Author.

American Psychiatric Association (APA). (1994). *Diagnostic and statistical manual of mental disorders: Fourth edition (DSM-IV)*. Washington, DC: Author.

American Psychiatric Association (APA). (2013). *Diagnostic and statistical manual of mental disorders, fifth edition (DSM-5)*. Washington, DC: Author.

Armour, C., Tsai, J., Durham, T. A., Charak, R., Biehn, T. L., Elhai, J. D., & Pietrzak, R. H. (2015). Dimensional structure of DSM-5 posttraumatic stress symptoms: Support for a hybrid anhedonia and externalizing behaviors model. *Journal of Psychiatric Research*, *61*, 106–113.

Babor, T. F., Higgins-Biddle, J. C., Saunders, J. B., Monteiro, M. G., & World Health Organization. (2001). AUDIT: The alcohol use disorders identification test: Guidelines for use in primary health care.

Bovin, M. J., Black, S. K., Rodriquez, P., Lunney, C. A., Kleiman, S. E., Weathers, F. W., Schnurr, P. P., Spira, J., Keane, T. M., & Marx, B. P. (revise and resubmit). Development and validation of a measure of PTSD-related functional impairment: The Inventory of Psychosocial Functioning.

Brewin, C. (2013). "I wouldn't start from here"—An alternative perspective on PTSD from the ICD-11: Comment on Friedman (2013). *Journal of Traumatic Stress*, *26*(5), 557–559.

Brewin, C. R., Lanius, R. A., Novac, A., Schnyder, U., & Galea, S. (2009). Reformulating PTSD for DSM-V: Life after criterion A. *Journal of Traumatic Stress*, *22*(5), 366–373.

Briere, J. (2001). *Detailed assessment of posttraumatic stress (DAPS)*. Odessa, FL: Psychological Assessment Resources.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290.

Cloitre, M., Garvert, D. W., Brewin, C. R., Bryant, R. A., & Maercker, A. (2013). Evidence for proposed ICD-11 PTSD and complex PTSD: A latent profile analysis. *European Journal of Psychotraumatology*, *4*, 20706.

Conigrave, K. M., Hall, W. D., & Saunders, J. B. (1995). The AUDIT questionnaire: Choosing a cut-off score. *Addiction*, *90*(10), 1349–1356.

IBM Corp, I. (2016). *IBM SPSS statistics for windows (version 24)*. Armonk, NY: IBM Cor.

(n.d.). Find comparative error, difference, significance of sample size. Retrieved from https://www.easycalculation.com/statistics/statistical-significance.php

First, M., Williams, J., Karg, R., & Spitzer, R. (2015). *Structured Clinical Interview for DMS-5, research version*. Arlington, VA: American Psychiatric Association.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1996). *Structured Clinical Interview for DSM-IV axis 1 disorders (SCID-I)*. Washington, DC: American Psychiatric Association.

Fullerton, C. S., Ursano, R. J., Epstein, R. S., Crowley, B., Vance, K., Kao, T.-C., … Baum, A. (2001). Gender differences in posttraumatic stress disorder after motor vehicle accidents. *American Journal of Psychiatry*, *158*(9), 1486–1491.

Green, B. (2003). Post-traumatic stress disorder: Symptom profiles in men and women. *Current Medical Research and Opinion*, *19*(3), 200–204.

Hafstad, G., Thoresen, S., Wentzel-Larsen, T., Maercker, A., & Dyb, G. (2017). PTSD or not PTSD? Comparing the proposed ICD-11 and the DSM-5 PTSD criteria among young survivors of the 2011 Norway attacks and their parents. *Psychological Medicine*, *47*(7), 1283–1291.

Hoge, C. W., Yehuda, R., Castro, C. A., McFarlane, A. C., Vermetten, E., Jetly, R., … Rauch, S. A. (2016). Unintended consequences of changing the definition of posttraumatic stress disorder in DSM-5: Critique and call for action. *JAMA Psychiatry*, *73*(7), 750–752.

Holowka, D. W., Marx, B. P., Kaloupek, D. G., & Keane, T. M. (2012). PTSD symptoms among male Vietnam veterans: Prevalence and associations

with diagnostic status. *Psychological Trauma: Theory, Research, Practice, and Policy*, 4(3), 285–292.

Keane, T. M., Caddell, J. M., & Taylor, K. L. (1988). Mississippi scale for combat-related posttraumatic stress disorder: Three studies in reliability and validity. *Journal of Consulting and Clinical Psychology*, 56(1), 85–90.

Keane, T. M., Rubin, A., Lachowicz, M., Brief, D., Enggasser, J. L., Roy, M., … Rosenbloom, D. (2014). Temporal stability of DSM–5 posttraumatic stress disorder criteria in a problem-drinking sample. *Psychological Assessment*, 26(4), 1138–1145.

Kraemer, H. C. (1992). *Evaluating medical tests: objective and quantitative guidelines*. Newbury Park, CA: Sage publications.

Mackinnon, A. (2000). A spreadsheet for the calculation of comprehensive statistics for the assessment of diagnostic tests and inter-rater agreement. *Computers in Biology and Medicine*, 30(3), 127–134.

Maercker, A., Brewin, C. R., Bryant, R. A., Cloitre, M., Reed, G. M., van Ommeren, M., … Llosa, A. E. (2013). Proposals for mental disorders specifically associated with stress in the International Classification of Diseases-11. *The Lancet*, 381(9878), 1683–1685.

Prins, A., Bovin, M. J., Smolenski, D. J., Marx, B. P., Kimerling, R., Jenkins-Guarnieri, M. A., … Tiet, Q. Q. (2016). The primary care PTSD screen for DSM-5 (PC-PTSD-5): development and evaluation within a veteran primary care sample. *Journal of General Internal Medicine*, 31(10), 1206–1211.

Prins, A., & Ouimette, P. (2004). The primary care PTSD screen (PC-PTSD): Development and operating characteristics. (vol 16, p. 257, 2003). *Primary Care Psychiatry*, 9(4), 151.

Ramchand, R., Schell, T. L., Karney, B. R., Osilla, K. C., Burns, R. M., & Caldarone, L. B. (2010). Disparate prevalence estimates of PTSD among service members who served in Iraq and Afghanistan: Possible explanations. *Journal of Traumatic Stress*, 23(1), 59–68.

Regier, D. A., Narrow, W. E., Clarke, D. E., Kraemer, H. C., Kuramoto, S. J., Kuhl, E. A., & Kupfer, D. J. (2013). DSM-5 field trials in the United States and Canada, part II: Test-retest reliability of selected categorical diagnoses. *American Journal of Psychiatry*, 170(1), 59–70.

Resick, P., Monson, C., & Chard, K. (2006). Cognitive processing therapy: Veteran/military version. *Clinical Psychology*, 74, 898–907.

Rodriguez, P., Holowka, D. W., & Marx, B. P. (2012). Assessment of posttraumatic stress disorder-related functional impairment: A review. *Journal of Rehabilitation Research and Development*, 49(5), 649–666.

Rosen, G. M., & Lilienfeld, S. O. (2008). Posttraumatic stress disorder: An empirical evaluation of core assumptions. *Clinical Psychology Review*, 28(5), 837–868.

Rosen, G. M., Lilienfeld, S. O., Frueh, B. C., McHugh, P. R., & Spitzer, R. L. (2010). Reflections on PTSD's future in DSM–V. *British Journal of Psychiatry*, 197(5), 343–344.

Rosen, G. M., Spitzer, R. L., & McHugh, P. R. (2008). Problems with the post-traumatic stress disorder diagnosis and its future in DSM–V. *British Journal of Psychiatry*, 192(1), 3–4.

Rosen, R. C., Marx, B. P., Maserejian, N. N., Holowka, D. W., Gates, M. A., Sleeper, L. A., … Keane, T. M. (2012). Project VALOR: Design and methods of a longitudinal registry of post-traumatic stress disorder (PTSD) in combat-exposed Veterans in the Afghanistan and Iraqi military theaters of operations. *International Journal of Methods in Psychiatric Research*, 21(1), 5–16.

Saunders, J. B., Aasland, O. G., Babor, T. F., De la Fuente, J. R., & Grant, M. (1993). Development of the alcohol use disorders identification test (AUDIT). WHO collaborative project on early detection of persons with harmful alcohol consumption-II. *Addiction*, 88, 791–791.

Spitzer, R. L., First, M. B., & Wakefield, J. C. (2007). Saving PTSD from itself in DSM-V. *Journal of Anxiety Disorders*, 21(2), 233–241.

Spitzer, R. L., Kroenke, K., & Williams, J. B. & Group, P. H. Q. P. C. S.(1999). Validation and utility of a self-report version of PRIME-MD: The PHQ primary care study. *Journal of the American Medical Association*, 282(18), 1737–1744.

Spitzer, R. L., Rosen, G. M., & Lilienfeld, S. O. (2008). Revisiting the Institute of Medicine report on the validity of posttraumatic stress disorder. *Comprehensive Psychiatry*, 49(4), 319–320.

Weathers, F. W., Blake, D. D., Schnurr, P. P., Kaloupek, D. G., Marx, B. P., & Keane, T. M. (2013). *Clinician-administered PTSD scale for DSM–5 (CAPS-5)*, Boston, MA: National Center for PTSD.

Zlotnick, C., Zimmerman, M., & Wolfsdorf, B. (2001). Gender differences in patients with posttraumatic stress disorder in a general psychiatric practice. *American Journal of Psychiatry*, 158, 1923–1925.