Received: 20 May 2022 | Revised: 22 December 2022 | Accepted: 26 January 2023

DOI: 10.1002/mpr.1963

WILEY

ORIGINAL ARTICLE

Differential symptom weighting in estimating empirical thresholds for underlying PTSD severity: Toward a "platinum" standard for diagnosis?

Antonio A. Morgan-López[1] | Lissette M. Saavedra[1] | Denise A. Hien[2] | Sonya B. Norman[3] | Skye S. Fitzpatrick[4] | Ai Ye[5,6] | Therese K. Killeen[7,8] | Lesia M. Ruglass[9] | Shannon M. Blakey[1] | Sudie E. Back[7,8]

Correspondence
Antonio A. Morgan-López, Community Health Research Division, RTI International, 3040 E. Cornwallis Road; Research Triangle Park, NC 27709, USA.
Email: amorganlopez@rti.org

Funding information
National Institute on Drug Abuse, Grant/Award Number: NIDA Clinical Trials Network Protocol 0015; National Institute of Alcohol Abuse and Alcoholism, Grant/Award Number: R01AA025853

Abstract

Objective: Symptom counts as the basis for Post-Traumatic Stress Disorder (PTSD) diagnoses in the DSM presume each symptom is equally reflective of underlying disorder severity. However, the "equal weight" assumption fails to fit PTSD symptom data when tested. The present study developed an enhanced PTSD diagnosis based on (a) a conventional PTSD diagnosis from a clinical interview and (b) an empirical classification of full PTSD that reflected the relative clinical weights of each symptom.

Method: Baseline structured interview data from Project Harmony ($N = 2658$) was used. An enhanced diagnosis for full PTSD was estimated using an empirical threshold from moderated nonlinear factor analysis (MNLFA) latent PTSD scale scores, in combination with a full conventional PTSD diagnosis based on interview data.

Results: One in 4 patients in the sample had a PTSD diagnosis that was inconsistent with their empirical PTSD grouping, such that the enhanced diagnostic standard reduced the diagnostic discrepancy rate by 20%. Veterans, and in particular female Veterans, were at greatest odds for discrepancy between their underlying PTSD severity and DSM diagnosis.

Conclusion: Psychometric methodologies that differentially weight symptoms can complement DSM criteria and may serve as a platform for symptom prioritization for diagnoses in future editions of DSM.

KEYWORDS
DSM, modeling, psychometrics, PTSD

The Consortium on Addiction, The Consortium on Addiction, Stress and Trauma (CAST) includes: Steven Batki, Malcolm Battersby, Matthew Boden, Deborah Brief, Christy Capone, Kathleen Chard, Joan Cook, Annette Crisanti, Erica Eaton, Thomas Ehring, Paul Emmelkamp, Edna Foa, Linda Frisman, Moira Haller, Deborah Kaysen, Shannon Kehle-Forbes, Asa Magnusson, Meghan McDevitt-Murphy, Mark McGovern, Katherine Mills, Lisa Najavits, David Oslin, Jessica Peirce, Ismene Petrakis, M. Zachary Rosenthal, Michael Saladin, Claudia Sannibale, Rebecca Schacht, Ingo Schaefer, Jeremiah Schumm, Tracy Simpson, Susan Sonne, Geraldine Tapia, Jessica Tripp, Debora Van Dam, Anka Vujanovic, and Caron Zlotnick.

1 | INTRODUCTION

Psychiatric disorders within the *Diagnostic and Statistical Manual of Mental Disorders* (DSM) are typically diagnosed by a simple count of symptoms. This is true for Post-Traumatic Stress Disorder (PTSD), which under DSM-5 requires (a) direct experiencing, witnessing or exposure to the consequences and aftermath of a qualifying traumatic event under Criterion A, (b) at least 1 of the 5 symptoms under the Intrusion criterion (Criterion B), (c) at least 1 of the 2 symptoms under the Avoidance criterion (Criterion C), (d) at least 2 of the 7 symptoms under the Negative Alterations of Cognitions and Mood criterion (Criterion D) and (e) at least 2 of the 6 symptoms under the Arousal and Reactivity criterion (Criterion E). Notably, there is no particular prioritization of any one symptom over any other symptom within each PTSD criterion, as each symptom is treated as though it has equal priority in reflecting underlying PTSD severity and diagnosis. However, it has been noted that symptom counts can give a distorted picture of underlying disorder severity, because the relative weight of each symptom in relation to the disorder is not taken into account (He et al., 2014; Morgan-López, Hien et al., 2022; Silverstein et al., 2020).

Concerns regarding the use of symptom counts to characterize a diagnosis, without prioritizing which symptoms are contributing to the diagnosis, parallel concerns about the use of total scores (or sum scores) to capture underlying PTSD severity. Most etiological and treatment outcome studies characterize PTSD severity by using raw total scores or symptom counts based on the sum of the categories on Likert-scaled items capturing the frequency and/or intensity of symptoms (e.g., Norman et al., 2019; Schäfer et al., 2019). However, it has been long understood that total scores can be problematic from a psychometric perspective (McNeish & Wolf, 2020) as well as a clinical perspective (Saavedra et al., 2022; Saavedra, Morgan-López, Hien, Back et al., 2021). Total scoring is primarily thought of as a "data management problem", often with little-to-no consideration for what is implied psychometrically by its use (Morgan-López, Hien et al., 2022; Morgan-López, Saavedra et al., 2022). For any latent variable model that is structured under the "equal weights" model (e.g., 1-parameter logistic item response theory (IRT), nonlinear factor analysis [NLFA] with equal factor loadings), the total score is a sufficient statistic for maximum likelihood estimation of the latent variable (Andrich, 1978, 2018; McNeish & Wolf, 2020). In practice however, it has been consistently demonstrated that PTSD symptoms fit to IRT or NLFA models require discrimination parameters/factor loadings that vary across symptoms (Contractor et al., 2018; Morgan-López, Saavedra et al., 2020; Ruglass et al., 2020; Saavedra, Morgan-López, Hien, Back, et al., 2021, 2022; Silverstein et al., 2020). The underlying clinical conclusion from this collection of advanced psychometric work is that some symptoms "matter" more than others (Bourne et al., 2013; Morgan-López, Hien et al., 2022).

A number of consequences have been noted when the "equal weights" assumption does not actually fit psychiatric symptom data that have implications for clinical decision-making (McNeish & Wolf, 2020; Sinharay & Haberman, 2014). In work specific to PTSD outcomes, it has been demonstrated that the use of total scores (compared to model-based factor analysis [FA]/IRT scores) can lead to non-trivial effect size distortion in the estimation of treatment outcomes (Morgan-López, Saavedra et al., 2020), mischaracterization of statistical inferences regarding differences in underlying PTSD severity across different racial/ethnic groups and across populations (e.g., incarcerated populations; Morgan-López, Hien et al., 2022; Ruglass et al., 2020), and discordance in judgments regarding whether individual patients had clinically significant improvement or deterioration (Saavedra, Morgan-López, Hien, Back et al., 2021, 2022).

Of particular relevance is the work described in Morgan-López, Killeen et al. (2020), where they developed an empirical classification approach that distinguishes patients with full PTSD and subthreshold PTSD that takes into account the differential weight of each symptom. The approach adapts the clinically significant change framework of Jacobson and Truax (1991), originally used to distinguish the "clinical" and "normative" groupings on a continuous psychiatric measure for treatment outcomes, but uses FA/IRT-based scale scores instead of total scores to distinguish between the clinical (full PTSD diagnosis group) and normative (subthreshold/no PTSD diagnosis group) scale score distributions. Greater separation of overlap between full PTSD and subthreshold PTSD/no PTSD scale score distributions has been observed when using FA/IRT-based scale scoring for PTSD severity compared to PTSD total scores/symptom counts (see Figure 2 in Saavedra, Morgan-López, Hien, Back et al., 2021). An empirical cutoff is estimated, in a manner similar to Jacobson and Truax (1991), by calculating the weighted midpoint between the full and subthreshold/no PTSD scale score distributions using the means and standard deviations of the scale scores for each group. Patients with a scale score above the cutoff are classified with an "empirical diagnosis" of full PTSD while patients who are below the cutoff are classified as not having full PTSD based on the empirical classification.

Demonstrating clear utility of the empirical classification approach, patients with a full DSM-IV PTSD diagnosis who also had a full PTSD empirical classification had significantly higher endorsement rates on 15 of 17 DSM-IV PTSD symptoms (and differed along several demographic variables) compared to patients with a full PTSD diagnosis *but a subthreshold PTSD empirical classification* (Morgan-López, Killeen et al., 2020). This suggests that there is substantial heterogeneity in the PTSD diagnosis– even among patients with a full PTSD diagnosis—and symptom counts (as opposed to weighting procedures such as FA/IRT) may obfuscate who has PTSD and how severe it is. The present study examines the utility of an enhanced diagnostic standard for PTSD, such that if the "gold" standard is a PTSD diagnosis based on (unweighted) within-criterion symptom counts from a clinical interview (e.g., Clinician-Administered PTSD Scale (CAPS); King et al., 1998), then a "platinum" standard diagnosis would be a conventional DSM PTSD diagnosis *plus* an empirical classification of full PTSD based on symptom weighting procedures (Morgan-López, Killeen et al., 2020). Anywhere from 25% to 50% of patients have shown discrepancies in both diagnostic grouping and

individual-level treatment outcome inference when comparing FA/IRT scores to raw scores/symptom counts (Jabraiylov et al., 2016; Morgan-López, Killeen et al., 2020; Morgan-López, Saavedra et al., 2022; Saavedra, Morgan-López, Hien, Back et al., 2021). Thus, it is critically important a) to assess the relative utility of this enhanced diagnostic standard and b) to understand the conditions and populations that may be most vulnerable to mischaracterization of their clinical status, in relation to their underlying PTSD severity after exposure to traumatic events.

2 | METHOD

2.1 | Participants

Baseline data were used from a subset of 25 studies (n = 2658) that are a part of an individual patient meta-analysis of behavioral and/or pharmacological treatments targeting PTSD comorbid with AODs known as *Project Harmony* (2022, in press; Hien et al., 2019, 2022, in press; Morgan-López, Hien et al., 2022; Saavedra, Morgan-López, Hien, López-Castro et al., 2021). These 25 studies are a subset of 36 studies in *Project Harmony* (N = 4046) that had item-level data on PTSD symptoms from semi-structured interview. Table 1 shows the studies that were included in the current analysis.

2.2 | Measures

PTSD items across all measures were recoded to a common standard at the item-level reflecting a binary (0/1) measure of presence/absence of the 16 DSM-IV/DSM-5 common symptoms and the 5 non-common symptoms, a process known in the integrative data analysis (IDA) literature as item harmonization (Bauer & Hussong, 2009).

TABLE 1 Study characteristics (N = 2658).

Study	N	Population type	Percent female	Percent full PTSD diagnosis	Measure
Back (2016)	27	Veteran	3.7	66.7	CAPS-IV
Back (2019)	81	Veteran	9.9	100.0	CAPS-IV
Boden (2012)	98	Veteran	0.0	86.7	CAPS-IV
Foa (2013)	165	Civilian/Veteran	34.5	72.1	PSS-I(-IV)
Hien (2004)	126	Civilian	100.0	82.5	CAPS-IV
Ruglass/Hien (2017)	95	Civilian	81.0	48.2	CAPS-IV
Hien (2009)	353	Civilian	100.0	80.1	CAPS-IV
Hien (2015)	113	Civilian	37.4	40.7	CAPS-IV
McDevitt-Murphy (2014)	68	Veteran	8.8	58.8	CAPS-IV
McGovern (2011)	53	Civilian	56.6	100.0	CAPS-IV
McGovern (2015)	284	Civilian	58.5	100.0	CAPS-IV
Mills (2012)	103	Civilian	62.1	95.2	CAPS-IV
Myers/Norman (2015)	41	Civilian	100.0	68.3	CAPS-IV
Norman (2019)	119	Veteran	10.1	86.6	CAPS-5
Haller/Norman (2016)	154	Veteran	11.7	64.3	CAPS-IV
Petrakis (2016)	96	Civilian	6.3	83.3	CAPS-IV
Petrakis (2020)	24	Veteran	8.4	29.1	CAPS-IV
Saladin (unpublished)	44	Civilian	50.0	93.1	CAPS-IV
Sannibale (2013)	62	Civilian	53.2	91.9	CAPS-IV
Schacht/Peirce (2017)	58	Civilian	79.3	100.0	CAPS-IV
Schafer (2019)	343	Civilian	100.0	63.3	PSS-I(-IV)
Sonne (unpublished)	25	Civilian	32.0	92.0	CAPS-IV
Vujanovic (2018)	53	Civilian	50.9	73.6	CAPS-5
Zlotnick (2003)	24	Incarcerated	100.0	87.5	CAPS-IV
Zlotnick (2009)	49	Incarcerated	100.0	81.6	CAPS-IV

Abbreviation: PTSD, Post-Traumatic Stress Disorder.

CAPS-IV Symptoms. The CAPS-IV (Blake et al., 1990) was used in 21 of the 25 studies in this analysis. The CAPS-IV assesses the frequency and intensity of the 17 DSM-IV PTSD symptoms respondents had experienced in the previous 30 days as well as to determine PTSD diagnostic status and symptom severity. The CAPS-IV has three symptom cluster subscales: Reexperiencing, Avoidance/Numbing, and Hyperarousal. A DSM-IV diagnosis of PTSD requires the presence of a Criterion A traumatic event, at least one reexperiencing symptom, three avoidance/numbing symptoms, and two hyperarousal symptoms. For harmonizing CAPS-IV items to a common metric relative to the other interview measures (CAPS-5, PSS-I-IV; see below), the convention for converting frequency and intensity items to binary DSM symptoms based on a symptom frequency rating of at least once in the previous month and a moderate or higher level of symptom intensity (known in the PTSD literature as "F1/I2") was used (20–21).

CAPS-5 Symptoms. The CAPS-5 (Weathers et al., 2018) was used in 2 of the 25 studies (Norman et al., 2019; Vujanovic et al., 2018) in this analysis. For the 20 items that capture severity of symptoms under DSM-5 Criteria B-E, harmonization to a common metric was done based on the item-level rule of a severity score of "2" (moderate) or higher based on DSM-5 PTSD criteria.

PSS-I(-IV) symptoms. The PSS-I (Foa et al., 1993) was used in 2 of the 25 studies (Foa et al., 2013; Schäfer et al., 2019) in this analysis. For the 17 items that capture severity of symptoms under DSM-IV Criteria B-D, harmonization to a common metric was done based on the item-level rule of a severity score of 2 ("2–4 times per week/somewhat") or higher based on DSM-IV PTSD criteria.

Predictors of Joint Empirical/Diagnostic Grouping. Predictors included the following demographic variables: age, gender, race/ethnicity, education level, marital status and population type (Civilian, Veteran, Currently Incarcerated). Two dummy variable indicators were included for the measure (CAPS-IV, CAPS-5, PSS-I-IV), with CAPS-IV as the reference measure. Other psychiatric predictors that were commonly available across datasets included: number of days of alcohol use in the past 30 days, any past 30-day cocaine use, any past 30-day opiate use, any past 30-day stimulant use, any past 30-day sedative use, any concomitant psychiatric medications, and current depression diagnosis. Descriptive statistics for predictors are shown in Table 2.

2.3 | Analysis procedures

PTSD Scale Score Estimation under Moderated Non-Linear Factor Analysis. Scale scores (set to a standard normal metric) were estimated under the moderated nonlinear factor analysis framework (MNLFA; Bauer, 2017) in Mplus (Muthén & Muthén, 1998-2017); the MNLFA model is ideal for estimating scale scores in the presence of differential item functioning (DIF). The model that best fit the harmonized PTSD symptoms, summarizing from Morgan-López, Hien et al. (2022), was a single factor, "2pl analog" MNLFA model with a) three symptoms showing no DIF across any predictor of DIF (i.e., psychologicalcues, negative beliefs, and horror/shame/guilt) and b) DIF substantial enough to lead to differential *test* functioning (i.e., systematic bias in scale scoring if DIF was not properly modeled) among African Americans and incarcerated populations. The MNLFA model was chosen because a) the general "2pl analog" model fit the data (CFI = 0.90, RMSEA = 0.052, 95% CI [0.048, 0.055]) while a "1pl analog" model with factor loadings constrained to equality failed to fit the data (CFI = 0.79, RMSEA = 0.072, 95% CI [0.068, 0.075]) and b) previous psychometric analyses suggested both item threshold and factor loading DIF was likely across multiple variables (e.g., Hoyt & Yeater, 2010; Jamison-Eddinger & McDevitt-Murphy, 2017; Ruglass et al., 2020).

Empirical Threshold for Distinguishing Full and Subthreshold PTSD. Next, an empirical threshold was estimated for differentiating MNLFA scale scores between participants with a full or subthreshold PTSD diagnosis using the means and standard deviations of the MNLFA scores (Jacobson & Truax, 1991):

$$\frac{(\sigma_{\text{Full PTSD}} x\ \text{Mean}_{\text{Subthreshold PTSD}}) + (\sigma_{\text{Subthreshold PTSD}} x\ \text{Mean}_{\text{Full PTSD}})}{(\sigma_{\text{Subthreshold PTSD}} + \sigma_{\text{Full PTSD}})} \quad (1)$$

Participants whose score was above the threshold would have an empirical full PTSD classification, while participants with a score below the threshold value would have an empirical subthreshold PTSD classification.

Comparison of the "Gold" and "Platinum" Standards for Diagnosis. The utility of adding a component to a full PTSD diagnosis that was dependent on both (a) a conventional diagnosis from a clinical interview (based on within-criterion symptom counts[1]) and (b) an empirical classification of full PTSD that was dependent on the relative weight of each symptom was examined; if (a) alone is considered a "Gold Standard" diagnosis, then we consider (a) and (b) together as a "Platinum Standard" for diagnosis. Using the empirical classifications from MNLFA and the diagnoses of Full PTSD and Subthreshold PTSD, four groupings were created: (1) a grouping where participants had both a Full PTSD diagnosis and a full PTSD empirical grouping (i.e., "Platinum Standard" diagnosis grouping), (2) two groups where the diagnosis and empirical classifications were discrepant (Full PTSD diagnosis/Subthreshold PTSD empirical, Subthreshold PTSD diagnosis/Full PTSD empirical) and (3) a grouping where participants had both a Subthreshold PTSD diagnosis and a Subthreshold PTSD empirical grouping.

Receiver Operating Curves (ROC) and sensitivity analysis for the two standards at various points along the latent PTSD MNLFA scale were used to assess differences in clinical utility for "gold" and "platinum" standard diagnoses. The predictors that were associated with higher probabilities of being in the diagnostic-discrepant groups were assessed using random effects multinomial logistic regression to assess which characteristics (i.e., populations, comorbidities) were most likely to lead to mischaracterization of participants' diagnostic/empirical PTSD grouping. Finally, to supplement validation of each diagnostic classification method as recommended by Jo et al. (2017), an assessment of the how each approach to diagnostic classification

TABLE 2 Descriptive statistics for predictors of diagnostic/empirical categorization groups.

Predictor	*N*	Mean/Proportion	SD	Min	Max
Age	2587	40.22	11.43	18	75
Male	2616	0.41	0.49	0	1
Hispanic	2623	0.07	0.26	0	1
White	2623	0.64	0.48	0	1
Black	2623	0.27	0.44	0	1
Other	2623	0.04	0.19	0	1
Asian	2623	0.01	0.09	0	1
High School or less	2011	0.50	0.50	0	1
Some college	1968	0.33	0.47	0	1
College	1968	0.16	0.36	0	1
Married	2225	0.18	0.39	0	1
Veteran	2601	0.36	0.48	0	1
Civilian	2601	0.62	0.49	0	1
Incarcerated	2658	0.03	0.16	0	1
Depression diagnosis	2273	0.57	0.51	0	1
CAPS-5	2658	0.06	0.25	0	1
PSS-I-IV	2658	0.19	0.39	0	1
Full PTSD diagnosis	2658	0.78	0.42	0	1
Days alcohol past 30 days	2496	9.96	10.87	0	30
Any cocaine past 30 days	1886	0.24	0.43	0	1
Any opiate past 30 days	1766	0.11	0.31	0	1
Any stimulant past 30 days	1744	0.10	0.30	0	1
Any sedative past 30 days	1522	0.10	0.30	0	1
Concomitant psychiatric Meds	1381	0.47	0.50	0	1

Abbreviation: PTSD, Post-Traumatic Stress Disorder.

of full PTSD (DSM diagnosis-only (DX), empirical classification-only (EC), DX + EC) predict PTSD treatment outcomes over time (Hien et al., in press).

3 | RESULTS

Empirical Threshold between Full and Subthreshold PTSD groups. Using Equation (1), the threshold dividing the Full PTSD group (mean = 0.22, SD = 0.85) and Subthreshold PTSD group (mean = –0.61, SD = 1.09) was –0.14. The subsequent diagnosis x empirical classification groupings are shown in Table 3.

TABLE 3 Counts of the four agreement groups.

Reference	Group 1	Group 2	Group 3
1373	694	195	395
(51.6%)	(26.0%)	(7.3%)	(14.9%)

Note: Reference Group: Diagnosis & Empirical scores agree that patient has full PTSD ("Platinum" Standard); Group 1: Diagnosis & Empirical Disagree: Diagnosis of full PTSD but empirical class says subthreshold PTSD; Group 2: Diagnosis of subthreshold PTSD but empirical class says full PTSD; Group 3: Diagnosis & Empirical scores agree that patient has subthreshold PTSD.

Abbreviation: PTSD, Post-Traumatic Stress Disorder.

3.1 | Diagnostic recovery and evaluation of the two standards

Next, we assessed improvements in clinical indicators of the "Platinum" standard diagnosis (Full PTSD diagnosis + Full PTSD empirical categorization) versus the "Gold" standard (Full PTSD diagnosis only). The ROC are shown in Figure 1, estimated based on logistic regression models with the PTSD MNLFA scale scores as the predictor (given the superior psychometric model fit compared to total symptom counts) and the diagnostic categories based on (a) the DSM diagnosis and (b) DSM diagnosis + empirical grouping as outcomes.

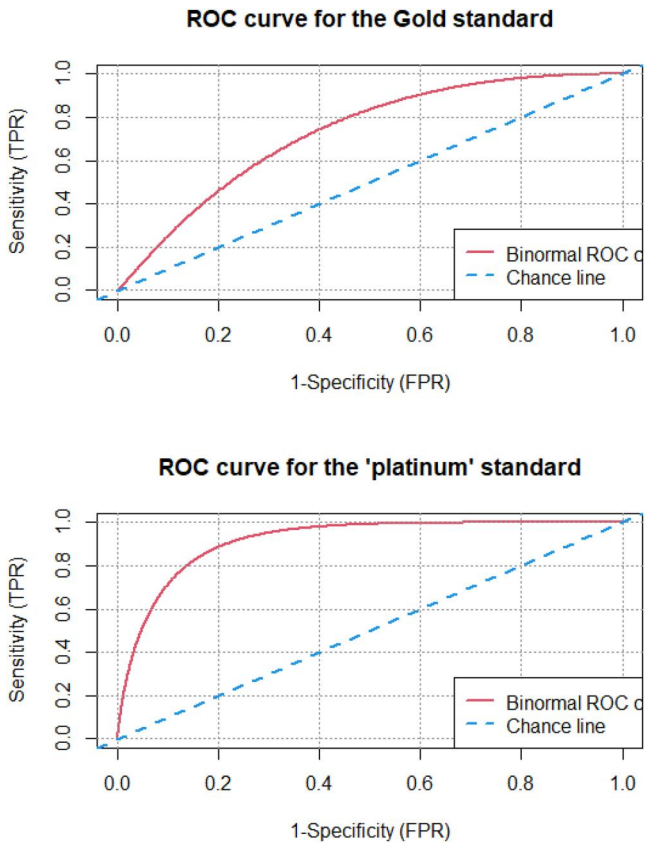

FIGURE 1 Receiver Operating Curves (ROC) Curve of Gold Standard versus Platinum Standard Diagnostic Categorizations. The Gold standard ROC curve represents 73% diagnostic concordance. The Platinum standard ROC curve represents 93% diagnostic concordance.

The area under the curve (AUC) value for the Platinum Standard diagnosis was 0.93 using the PTSD MNLFA score as the model predictor. Using the PTSD MNLFA score as the model predictor for a conventional/Gold Standard diagnosis, the AUC value was 0.73, suggesting that greater than 1 in 4 patients (27%) have a diagnosis that is inconsistent with their empirical PTSD grouping. Further, the difference between the two methods is particularly pronounced when balancing specificity and sensitivity within the Figure for example, when specificity is 80%, the sensitivity of the "Gold Standard" was just around 48% while that of the "Platinum Standard" was around 89%.

3.2 | Variation in diagnostic/empirical grouping proportions by covariates

Given the marginal percentages of the diagnostic/empirical PTSD groupings as shown in Table 3, the extent to which key covariates were predictive of differences across classifications were examined using random effects multinomial logistic regression, with a focus on covariates that increased the probabilities of being in the diagnosis-discrepant groupings. However, due covariate missingness, 50 multiply imputed datasets were generated under a random intercept multilevel structure (adjusting for within-study clustering) using the R package "mice" (Van Buuren & Groothuis-Oudshoorn, 2011); results are presented as pooled estimates and standard errors across all 50 imputed datasets.

Full Diagnosis/Subthreshold Empirical versus "Platinum Standard" group. Compared to the total sample, older patients (odds = 1.01 [CI = 1.001, 1.02]), patients reporting past month cocaine use (odds = 1.32 [CI = 1.01, 1.73]) and patients assessed using the PSS-I (odds = 1.90 [CI = 1.36, 2.64]) had a significantly greater odds of receiving a Full PTSD diagnosis but being empirically classified as subthreshold. Participants who had a significantly lower odds of this type of diagnostic discrepancy included men (odds = 0.72 [CI = 0.55, 0.93]), patients with a comorbid depression diagnosis (odds = 0.66 [CI = 0.54, 0.81]) and patients assessed using the CAPS-5 (odds = 0.19 [CI = 0.06, 0.64]).

Subthreshold Diagnosis/Full Empirical versus "Platinum Standard" group. Groups that had a significantly greater odds of receiving a Subthreshold PTSD diagnosis but being empirically classified as Full PTSD included veterans (odds = 2.46 [CI = 1.60, 3.79]) and married participants (odds = 1.76 [CI = 1.19, 2.60]). Participants who were less likely to be in this diagnostic-discrepant grouping included non-veteran men (odds = 0.48 [CI = 0.31, 0.73]), patients with a comorbid depression diagnosis (odds = 0.46 [CI = 0.33, 0.65]), patients entering treatment on concomitant non-study medications (odds = 0.55 [CI = 0.37, 0.82]) and patients assessed using the PSS-I (odds = 0.19 [CI = 0.06, 0.64]).

Full versus Subthreshold PTSD: Treatment Outcome Differences by Method. Assessment of further validation of the approaches to PTSD classification was done by estimation of PTSD treatment outcomes from baseline through end-of-treatment. Separate three-level multilevel models (repeated measures clustered within patients, patients clustered within studies) were fit to PTSD treatment outcome data in *Project Harmony* using PTSD scale score estimates jointly modeled under MNLFA from harmonized self-report and clinical interview data (see Supplemental Material for Hien et al., in press for details); the focal estimates were diagnostic group × Time interactions, which would denote differences in average treatment trajectories between full PTSD and subthreshold PTSD patients. The literature has suggested, though not consistently, that greater baseline PTSD severity may be associated with worse PTSD outcomes (Barawi et al., 2020).

Findings differed substantially across methods of diagnostic categorization. Using the DSM diagnosis, full and subthreshold PTSD outcome trajectories were nearly indistinguishable on average (Dx × Time interaction = −0.194 (0.103), $t = -1.88$, $p = 0.06$, $d = -0.24$). However, using the empirical grouping that is differentially weighted by each symptom, patients empirically classified as having full PTSD were significantly (and meaningfully based on effect size) worse off over time on average than their empirical subthreshold PTSD counterparts (EC × Time interaction = 0.760 (0.075), $t = 10.07$, $p < 0.001$, $d = 0.99$); similar findings were observed for the

Platinum Standard treatment outcome validation analysis (Platinum × Time interaction = 0.721 (0.082), $t = 8.78$, $p < 0.001$, $d = 0.94$).

4 | DISCUSSION

Using a subset of 25 studies integrated under *Project Harmony*, the present study evaluated an enhanced diagnostic standard for Full PTSD, whereby patients meet both a conventional DSM PTSD diagnosis and an empirical classification of Full PTSD based on symptom weighting procedures. In this sample, the marginal proportion of diagnostic agreement was 66.5%, suggesting that 1 in 3 patients had a discrepancy between their original PTSD diagnosis and their empirical PTSD grouping, an even larger diagnostic discrepancy than that observed in Morgan-López, Killeen et al. (2020). Given this discrepancy, it is tempting to ask the question "Which is correct: the DSM diagnosis or the empirical grouping?" Recall that, between the empirically estimated scale scores under (MNL)FA and a much more restrictive model assumed by using within-criteria symptom counts (i.e., the basis for diagnosis), the former model fit the data well and the latter model fit the data poorly. Thus, symptom counts form a biased basis, based on psychometric model misfit, for measuring underlying PTSD severity and arguably a biased basis from which to make a categorical decision regarding a PTSD diagnosis (Sinharay & Haberman, 2014). To many clinicians this will come as no surprise, as it is both clinically intuitive that "not all symptoms matter equally" and has been shown in multiple factor analytic and IRT-based psychometric studies (Contractor et al., 2018; Saavedra, Morgan-López, Hien, Back, et al., 2021; Silverstein et al., 2020).

Given the noted psychometric limitations of the DSM diagnosis itself, the focus in this article was on augmenting the diagnostic standard by requiring both the DSM Full PTSD diagnosis and an empirical grouping of Full PTSD for an enhanced or "Platinum Standard" diagnosis. Using the enhanced diagnostic standard increased the model-based ROC/AUC from 0.73 to 0.93 compared to a Full PTSD diagnosis only, reducing the estimated diagnostic discrepancy rate by 20%. This result is critical, because without attending to differences in the relative weight of PTSD symptoms, a non-trivial proportion of undiagnosed-but-high-severity patients may be at-risk for missing necessary services and, conversely, patients with a diagnosis whose underlying severity is not as high might otherwise be referred to more appropriate services. This enhanced diagnostic standard has the potential to be both more accurate and more nuanced, opening the door for research on PTSD treatment that is more personalized and tailored based on specificity in the symptoms that contribute most toward a patient's underlying severity.

A key population that was more likely to be in the "underdiagnosed" discrepancy group were veterans, who were 2.4 times more likely than other mutually exclusive populations (non-veteran civilians, incarcerated populations) to be subthreshold with regard to their PTSD diagnosis but empirically estimated to have Full PTSD, consistent with other work detailing risk for underdiagnosis of PTSD in veteran populations (Fisher, 2014). This finding is underscored by the higher marginal percentages for diagnostic discrepancy for female veterans (42%) versus male veterans (34%). Women veterans represent ~10% of patients seen in VA care clinically and in research (Department of Veterans Affairs, 2017). It is possible that women veterans on average differ in how their PTSD symptoms present because of greater exposure to previous trauma, to sexual harassment and assault, and to stressful homecoming experiences (Street et al., 2009). Further, this type of underdiagnosis is in contrast to non-veteran men who have lower odds of diagnostic discrepancy (marginal percentage of 23%; see also Table 4). More research is necessary to understand why veterans, particularly female veterans, are likely to be under-diagnosed with PTSD via conventional diagnosis in contrast to our proposed "Platinum" standard. One key population that was more likely to be "overdiagnosed" (i.e., Full PTSD diagnosis/subthreshold PTSD empirical classification) was patients with comorbid cocaine use. It is possible that patients may attribute the effects of cocaine use (but not other substances, such as alcohol use; Waldrop et al., 2007) to PTSD, resulting in higher PTSD symptom ratings by assessors. Indeed, one study found that the presence (compared to the absence) of cocaine use disorder was associated with significantly higher arousal/reactivity symptoms of PTSD (Dworkin et al., 2018). These authors suggested that certain experiences, such as such as hypervigilance and irritability and angry outbursts, could be either cocaine- or trauma-related. Accordingly, our findings underscore the importance of careful assessment of PTSD in the presence of concurrent cocaine use.

Our findings also suggest that assessment under DSM-5 reduces the odds of diagnostic discrepancy (see Table 4), with the marginal percentage for diagnostic discrepancy of 26% under CAPS-5 compared to 34% for CAPS-IV/PSS-I-IV. This is likely due to a larger influence of two DSM-5 symptoms in particular: in MNLFA scale score estimation, Negative Beliefs had the second largest factor loading and Blame of Self or Others had the second largest item threshold (Morgan-López, Hien et al., 2022), suggesting that these two "high information" symptoms contributed to greater precision in estimating the scale score and likely have greater impact on diagnosis from the two studies from which DSM-5 symptoms were available (Norman et al., 2019; Vujanovic et al., 2018). The additional information that these symptoms provide speaks to one limitation of the study (i.e., high proportion of DSM-IV studies), where there will be a need to extend this methodology to IDA contexts once a critical mass of DSM-5 PTSD/AOD studies are available. Relatedly, a second limitation of this study is that analyses were conducted on a set of studies of adults with comorbid PTSD/AOD. Given other work showing that people with PTSD/AOD, relative to PTSD only, have greater severity across numerous clinical variables, including PTSD symptom count (Simpson et al., 2019), we cannot be certain that observed "Gold" versus "Platinum" standard discrepancies would have emerged in an IDA study of adults with PTSD only; calls have been made recently for such IDA studies with "pure" PTSD samples

TABLE 4 Random effects multinomial logistic regression estimates.

Status	Predictor	Odds Ratio	Lower CL	Upper CL
Full PTSD Diagnosis/Subthreshold PTSD empirical grouping (v. Full/Full)	Age	1.0101	1.0013	1.0190
	Hispanic (v. Non-Hispanic)	1.2076	0.8295	1.7579
	Male	0.7171	0.5503	0.9345
	Black (v. White)	1.0373	0.8127	1.3240
	Other (v. White)	1.2274	0.7445	2.0235
	Asian (v. White)	0.9987	0.2999	3.3262
	Some college (v. High School)	1.0049	0.7957	1.2691
	College (v. High School)	0.8203	0.5968	1.1275
	Married	1.1341	0.8690	1.4800
	Veteran (v. Civilian)	1.1245	0.8593	1.4717
	CAPS-5 (v. CAPS-IV)	0.1902	0.0559	0.6471
	veteran*caps5	3.0345	0.8109	11.3551
	Incarcerated (v. Civilian)	0.2932	0.0686	1.2525
	Comorbid depression	0.6613	0.5389	0.8115
	Non-study medications	0.8329	0.6550	1.0592
	PSS-I-IV (v. CAPS-IV)	1.9005	1.3639	2.6483
	Number of days alcohol in past month	1.0057	0.9960	1.0154
	Any past month cocaine use	1.3219	1.0110	1.7285
	Any past month opiate use	1.0813	0.7236	1.6160
	Any past month stimulant use	0.7202	0.4722	1.0986
	Any past month sedative use	0.9433	0.6228	1.4288
Subthreshold PTSD Diagnosis/Full PTSD empirical grouping (v. Full/Full)	Age	1.0010	0.9862	1.0161
	Hispanic (v. Non-Hispanic)	0.7014	0.3659	1.3445
	Male	0.4752	0.3084	0.7321
	Black (v. White)	0.9188	0.6204	1.3608
	Other (v. White)	1.7382	0.8657	3.4903
	Asian (v. White)	2.2004	0.6030	8.0286
	Some college (v. High School)	1.2107	0.8323	1.7611
	College (v. High School)	0.8888	0.5488	1.4397
	Married	1.7635	1.1957	2.6011
	Veteran (v. Civilian)	2.4666	1.6041	3.7930
	CAPS-5 (v. CAPS-IV)	0.8579	0.2715	2.7105
	veteran*caps5	0.5950	0.1552	2.2808
	Incarcerated (v. Civilian)	0.1230	0.0054	2.8244
	Comorbid depression	0.4600	0.3268	0.6474
	Non-study medications	0.5537	0.3728	0.8224
	PSS-I-IV (v. CAPS-IV)	0.0906	0.0238	0.3455
	Number of days alcohol in past month	0.9992	0.9813	1.0175
	Any past month cocaine use	1.0762	0.6992	1.6564
	Any past month opiate use	0.8457	0.4391	1.6287
	Any past month stimulant use	0.9143	0.4964	1.6842
	Any past month sedative use	0.6763	0.3040	1.5045

Abbreviation: PTSD, Post-Traumatic Stress Disorder.

(Morgan-López, Hien et al., 2022). However, the high prevalence and unique impairments reported in civilian and military samples of patients with comorbid PTSD/AOD is an important area of focus (Norman et al., 2019). A third limitation is, although most studies utilized the CAPS, which is a gold standard measure of PTSD, "real world" clinical practice may not utilize structured interviews in determining PTSD diagnoses. Further work is needed to investigate the discrepancies between PTSD diagnoses based on unstructured clinical interviews and empirical classifications to inform assessment in non-research settings.

Although convenient and practical, the use of total scores or symptom counts impose a psychometric model where each symptom of a disorder is equally weighted in its relation to underlying disorder severity. The current study shows that this one size fits all approach can lead to over- *and* under-diagnosing of PTSD in specific subgroups, potentially contributing to disparities in who is eligible for services and treatment for PTSD (Spoont et al., 2009, 2017). This study adds to growing evidence of the utility of using FA/IRT scale score estimation in concert with normative threshold approaches for categorical decision-making (Morgan-López, Killeen et al., 2020), such that diagnostic decision-making can a) take into account the relative weighting of symptoms and the extent to which FA/IRT item parameters differ across populations and b) potentially yield a different diagnostic decision than historical and current (within-criterion) symptom count approaches to diagnosis within the DSM. By empirically identifying potential "red flag" symptoms (relative to less severe "yellow flag" symptoms) that can enhance diagnostic precision, it is possible that this more nuanced and accurate approach can help the trauma field achieve the goal of better understanding and treating the after-effects of trauma in all people.

AFFILLATIONS

[1]Community Health Research Division, RTI International, Research Triangle Park, North CarolinaUSA

[2]Center of Alcohol & Substance Use Studies, Rutgers University–New Brunswick, Piscataway, New JerseyUSA

[3]Department of Psychiatry, University of California, San Diego, VirginiaUSA

[4]Department of Psychology, York University, Toronto, OntarioCanada

[5]Department of Psychology & Neuroscience, L.L. Thurstone Psychometric Laboratory, UNC-Chapel Hill, Chapel Hill, North CarolinaUSA

[6]Department Psychologie, Ludwig-Maximilians-Universität, Munich, Germany

[7]Department of Psychiatry and Behavioral Sciences, Medical University of South Carolina, Charleston, South CarolinaUSA

[8]Ralph H. Johnson VA Medical Center, Charleston, South CarolinaUSA

[9]Department of Psychology, City College of New York, New York, New YorkUSA

ACKNOWLEDGMENTS

This work was funded by grants from by the National Institute on Drug Abuse (NIDA Clinical Trials Network Protocol 0015, Hien, PI) and the National Institute of Alcohol Abuse and Alcoholism (R01AA025853; Hien & Morgan-López, MPIs).

DATA AVAILABILITY STATEMENT

The data that support the findings (with removal of data from studies that do not allow data sharing beyond Project Harmony Investigators) will be available in the NIAAA Data Archive following a 12-month embargo from the completion date of the granting project to allow for dissemination of research findings.

ORCID

Antonio A. Morgan-López https://orcid.org/0000-0003-4706-9964

ENDNOTE

[1] Specific to the edition of the DSM used in each study.

REFERENCES

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*(4), 561–573. https://doi.org/10.1007/bf02293814

Andrich, D. (2018). Rasch rating-scale model. In *Handbook of item response theory* (pp. 75–94). Chapman and Hall/CRC.

Back, S. E., Killeen, T., Badour, C. L., Flanagan, J. C., Allan, N. P., Santa Ana, E., Lozano, B., Korte, K. J., Foa, E. B., & Brady, K. T. (2019). Concurrent treatment of substance use disorders and PTSD using prolonged exposure: A randomized clinical trial in military veterans. *Addictive Behaviors*, *90*, 369–377.

Back, S. E., McCauley, J. L., Korte, K. J., Gros, D. F., Leavitt, V., Gray, K. M., Hamner, M. B., DeSantis, S. M., Malcolm, R., Brady, K. T., & Kalivas, P. W. (2016). A double-blind, randomized, controlled pilot trial ofN-acetylcysteine in veterans with posttraumatic stress disorder and substanceuse disorders. *The Journal of Clinical Psychiatry*, *77*(11), 1439–1446.

Barawi, K. S., Lewis, C., Simon, N., & Bisson, J. I. (2020). A systematic review of factors associated with outcome of psychological treatments for post-traumatic stress disorder. *European Journal of Psychotraumatology*, *11*(1), 1774240. https://doi.org/10.1080/20008198.2020.1774240

Bauer, D. J. (2017). A more general model for testing measurement invariance and differential item functioning. *Psychological Methods*, *22*(3), 507–526. https://doi.org/10.1037/met0000077

Bauer, D. J., & Hussong, A. M. (2009). Psychometric approaches for developing commensurate measures across independent studies: Traditional and new models. *Psychological Methods*, *14*(4), 101–125. https://doi.org/10.1037/a0017642

Blake, D., Weathers, F., Nagy, L., Kaloupek, D., Klauminzer, G., Charney, D., & Keane, T. (1990). Clinician-administered PTSD scale (CAPS) (Vol. *7*).

Boden, M. T., Kimerling, R., Jacobs-Lentz, J., Bowman, D., Weaver, C., Carney, D., Walser, R., & Trafton, J. A. (2012). Seeking Safety treatment for male veterans with a substance use disorder and posttraumatic stress disorder symptomatology. *Addiction*, *107*(3), 578–586.

Bourne, C., Mackay, C. E., & Holmes, E. A. (2013). The neural basis of flashback formation: The impact of viewing trauma. *Psychological Medicine*, *43*(7), 1521–1532. https://doi.org/10.1017/S0033291712002358

Contractor, A. A., Caldas, S. V., Dolan, M., Lagdon, S., & Armour, C. (2018). PTSD's factor structure and measurement invariance across subgroups with differing count of trauma types. *Psychiatry Research*, *264*, 76–84. https://doi.org/10.1016/j.psychres.2018.03.065

Department of Veterans Affairs. (2017). *Women veterans report: The past, present, and future of women veterans*. National Center for Veterans Analysis and Statistics.

Dworkin, E. R., Wanklyn, S., Stasiewicz, P. R., & Coffey, S. F. (2018). PTSD symptom presentation among people with alcohol and drug use disorders: Comparisons by substance of abuse. *Addictive Behaviors*, *76*, 188–194. https://doi.org/10.1016/j.addbeh.2017.08.019

Fisher, M. P. (2014). PTSD in the US military, and the politics of prevalence. *Social Science & Medicine*, *115*, 1–9. https://doi.org/10.1016/j.socscimed.2014.05.051

Foa, E. B., Yusko, D. A., McLean, C. P., Suvak, M. K., Bux, D. A., Oslin, D., O'Brien, C. P., Imms, P., Riggs, D. S., & Volpicelli, J. (2013). Concurrent naltrexone and prolonged exposure therapy for patients with comorbid alcohol dependence and PTSD: A randomized clinical trial. *JAMA*, *310*, 488–495.

Foa, E. B., Yusko, D. A., McLean, C. P., Suvak, M. K., Bux, D. A., Oslin, D., O'Brien, C. P., Imms, P., Riggs, D. S., & Volpicelli, J. (2013). Concurrent naltrexone and prolonged exposure therapy for patients with comorbid alcohol dependence and PTSD: A randomized clinical trial. *JAMA*, *310*(5), 488–495. https://doi.org/10.1001/jama.2013.8268

Foa, E., Riggs, D. S., Dancu, C. V., Constance, V., & Rothbaum, B. O. (1993). Reliability and validity of a brief instrument for assessing post-traumatic stress disorder. *Journal of Traumatic Stress*, *6*(4), 459–473. https://doi.org/10.1002/jts.2490060405

Haller, M., Norman, S. B., Cummins, K., Trim, R. S., Xu, X., Cui, R., Allard, C. B., Brown, S. A., & Tate, S. R. (2016). Integrated cognitive behavioral therapy versus cognitive processing therapy for adults with depression, substance use disorder, and trauma. *Journal of Substance Abuse Treatment*, *62*, 38–48.

He, Q., Glas, C. A., & Veldkamp, B. P. (2014). Assessing impact of differential symptom functioning on post-traumatic stress disorder (PTSD) diagnosis. *International Journal of Methods in Psychiatric Research*, *23*(2), 131–141. https://doi.org/10.1002/mpr.1417

Hien, D. A., Cohen, L. R., Miele, G. M., Litt, L. C., & Capstick, C. (2004). Promising treatments for women with comorbid PTSD and substance use disorders. *American Journal of Psychiatry*, *161*(8), 1426–1432.

Hien, D. A., Fitzpatrick, S., Saavedra, L. M., Ebrahimi, C. T., Norman, S. B., Tripp, J., Ruglass, L. M., Lopez-Castro, T., Killeen, T. K., Back, S. E., Morgan-Lopez, A. A., & Morgan-López, A. A. (2022). What's in a name? A data-driven method to identify optimal psychotherapy classifications to advance treatment research on co-occurring PTSD and substance use disorders. *European Journal of Psychotraumatology*, *13*(1), 2001191. https://doi.org/10.1080/20008198.2021.2001191

Hien, D. A., Levin, F. R., Ruglass, L. M., López-Castro, T., Papini, S., Hu, M.-C., Cohen, L. R., & Herron, A. (2015). Combining seeking safety with sertraline for PTSD and alcohol use disorders: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *83*(2), 359–369. https://doi.org/10.1037/a0038719

Hien, D. A., Morgan-López, A. A., Ruglass, L. M., Saavedra, L. M., Fitzpatrick, S., Back, S. E., Killeen, T. K., & Norman, S. B. (2019). Project Harmony: A systematic review and meta-analysis of individual patient data of behavioral and pharmacologic trials for comorbid posttraumatic stress, alcohol, and other drug use disorders. Retrieved from https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42019146678

Hien, D. A., Morgan-López, A. A., Saavedra, L. M., Ruglass, L. M., Ye, A., López-Castro, T., Fitzpatrick, S., Killeen, T. K., Norman, S. B., Ebrahimi, C. T. & Back, S. E. (2022). Project Harmony: A meta-analysis with individual patient data of behavioral and pharmacologic trials for comorbid posttraumatic stress, alcohol and other drug use disorders. *American Journal of Psychiatry*, *180*, 155–166. https://doi.org/10.1176/appi.ajp.22010071

Hien, D. A., Wells, E. A., Jiang, H., Suarez-Morales, L., Campbell, A. N., Cohen, L. R., Miele, G. M., Killeen, T., Brigham, G. S., Zhang, Y., Hansen, C., Hodgkins, C., Hatch-Maillette, M., Brown, C., Kulaga, A., Kristman-Valente, A., Chu, M., Sage, R., Robinson, J. A., … Nunes, E. V. (2009). Multisite randomized trial of behavioral interventions for women with co-occurring PTSD and substance use disorders. *Journal of consulting and clinical psychology*, *77*(4), 607.

Hoyt, T., & Yeater, E. A. (2010). Comparison of posttraumatic stress disorder symptom structure models in Hispanic and White college students. *Psychological Trauma: Theory, Research, Practice, and Policy*, *2*(1), 19–30. https://doi.org/10.1037/a0018745

Jabrayilov, R., Emons, W. H. M., & Sijtsma, K. (2016). Comparison of classical test theory and item response theory in individual change assessment. *Applied Psychological Measurement*, *40*(8), 559–572. https://doi.org/10.1177/2F0146621616664046

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*(1), 12–19. https://doi.org/10.1037/0022-006x.59.1.12

Jamison-Eddinger, J. R., & McDevitt-Murphy, M. E. (2017). A confirmatory factor analysis of the PTSD Checklist 5 in veteran and college student samples. *Psychiatry Research*, *255*, 219–224. https://doi.org/10.1016/j.psychres.2017.05.035

Jo, B., Findling, R. L., Wang, C. P., Hastie, T. J., Youngstrom, E. A., Arnold, L. E., Fristad, M. A., & Horwitz, S. M. (2017). Targeted use of growth mixture modeling: A learning perspective. *Statistics in Medicine*, *36*(4), 671–686. https://doi.org/10.1002/sim.7152

King, D. W., Leskin, G. A., King, L. A., & Weathers, F. W. (1998). Confirmatory factor analysis of the Clinician-Administered PTSD Scale: Evidence for the dimensionality of posttraumatic stress disorder. *Psychological Assessment*, *10*(2), 90–96. https://doi.org/10.1037/1040-3590.10.2.90

McDevitt-Murphy, M. E., Murphy, J. G., Williams, J. L., Monahan, C. J., Bracken-Minor, K. L., & Fields, J. A. (2014). Randomized controlled trial of two brief alcohol interventions for OEF/OIF veterans. *Journal of Consulting and Clinical Psychology*, *82*(4), 562–568.

McGovern, M. P., Lambert-Harris, C., Alterman, A. I., Xie, H., & Meier, A. (2011). A randomized controlled trial comparing integrated cognitive behavioral therapy versus individual addiction counseling for co-occurring substance use and posttraumatic stress disorders. *Journal of Dual Diagnosis*, *7*(4), 207–227.

McGovern, M. P., Lambert-Harris, C., Xie, H., Meier, A., McLeman, B., & Saunders, E. (2015). A randomized controlled trial of treatments for co-occurring substanceuse disorders and post-traumatic stress disorder. *Addiction*, *110*, 1194–1204.

McNeish, D., & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior Research Methods*, *52*(6), 2287–2305. https://doi.org/10.3758/s13428-020-01398-0

Mills, K. L., Teesson, M., Back, S. E., Brady, K. T., Baker, A. L., Hopwood, S., Sannibale, C., Barrett, E. L., Merz, S., Rosenfeld, J., & Ewer, P. L. (2012). Integrated exposure-based therapy for co-occurring posttraumatic stress disorder and substance dependence: A randomized controlled trial. *Jama*, *308*(7), 690–699. https://doi.org/10.1001/jama.2012.9071

Morgan-Lopez, A. A., Saavedra, L. M., Hien, D. A., Killeen, T. K., Back, S. E., Ruglass, L. M., Fitzpatrick, S., Lopez-Castro, T., & Patock-Peckham, J. A. (2020). Estimation of equable scale scores and treatment outcomes from patient- and clinician-reported PTSD measures using item response theory calibration. *Psychological Assessment*, *32*(4), 321–335. https://doi.org/10.1037/pas0000789

Morgan-López, A. A., Hien, D. A., Saraiya, T. C., Saavedra, L. M., Norman, S. B., & Killeen, T. K. (2022). amp; Consortium on Addiction, Stress and Trauma (CAST)Estimating posttraumatic stress disorder severity in the presence of differential item functioning across populations, comorbidities, and interview measures: Introduction to Project Harmony. *Journal of Traumatic Stress*. https://doi.org/10.1002/jts.22800

Morgan-López, A. A., Killeen, T. K., Saavedra, L. M., Hien, D. A., Fitzpatrick, S., Ruglass, L. M., & Back, S. E. (2020). Crossover between diagnostic and empirical categorizations of full and subthreshold PTSD. *Journal*

of Affective Disorders, *274*(4), 832–840. https://doi.org/10.1016/j.jad.2020.05.031

Morgan-López, A. A., Saavedra, L. M., Ramirez, D. D., Smith, L. M., & Yaros, A. C. (2022). Adapting the multilevel model for estimation of the reliable change index (RCI) with multiple timepoints and multiple sources of error. *International Journal of Methods in Psychiatric Research*, *31*(2), e1906. https://doi.org/10.1002/mpr.1906

Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide* (8th ed.). Muthén & Muthén.

Myers, U. S., Browne, K. C., & Norman, S. B. (2015). Treatment engagement: Female survivors of intimate partner violence in treatment for PTSD and alcohol use disorder. *Journal of Dual Diagnosis*, *11*(3–4), 238–247.

Norman, S. B., Trim, R., Haller, M., Davis, B. C., Myers, U. S., Colvonen, P. J., Blanes, E., Lyons, R., Siegel, E. Y., Angkaw, A. C., Norman, G. J., & Mayes, T. (2019). Efficacy of integrated exposure therapy vs integrated coping skills therapy for comorbid posttraumatic stress disorder and alcohol use disorder: A randomized clinical trial. *JAMA Psychiatry*, *76*(8), 791–799. https://doi.org/10.1001/jamapsychiatry.2019.0638

Norman, S. B., Trim, R., Haller, M., Davis, B. C., Myers, U. S., Colvonen, P. J., Blanes, E., Lyons, R., Siegel, E. Y., Angkaw, A. C., Norman, G. J., & Mayes, T. (2019). Efficacy of integrated exposure therapy vs integrated coping skills therapy for comorbid posttraumatic stress disorder and alcohol use disorder: A randomized clinical trial. *JAMA Psychiatry*, *76*, 791–799.

Petrakis, I. L., Desai, N., Gueorguieva, R., Arias, A., O'Brien, E., Jane, J. S., Sevarino, K., Southwick, S., & Ralevski, E. (2016). Prazosin for veterans with posttraumatic stress disorder and comorbid alcohol dependence: A clinical trial. *Alcoholism: Clinical and Experimental Research*, *40*(1), 178–186.

Petrakis, I., Ralevski, E., Arias, A. J., DeNegre, D., Newcomb, J., Gianoli, M., McCarthy, E., Meshberg-Cohen, S., & Yoon, G. (2020). Zonisamide as an adjunctive treatment to cognitive processing therapy for veterans with posttraumatic stress disorder and comorbid alcohol use disorder: A pilot study. *The American Journal on Addictions*, *29*(6), 515–524.

Ruglass, L. M., Lopez-Castro, T., Papini, S., Killeen, T., Back, S. E., & Hien, D. A. (2017). Concurrent treatment with prolonged exposure for co-occurring full or subthreshold posttraumatic stress disorder and substance use disorders: Arandomized clinical trial. *Psychotherapy and Psychosomatics*, *86*(3), 150–161.

Ruglass, L. M., Morgan-López, A. A., Saavedra, L. M., Hien, D. A., Fitzpatrick, S., Killeen, T. K., Back, S. E., & López-Castro, T. (2020). Measurement nonequivalence of the Clinician-Administered PTSD Scale by race/ethnicity: Implications for quantifying posttraumatic stress disorder severity. *Psychological Assessment*, *32*(11), 1015–1032. https://doi.org/10.1037/pas0000943

Saavedra, L. M., Morgan-López, A. A., Back, S. E., Patel, S. V., Hien, D. A., Killeen, T. K., Norman, S. B., Fitzpatrick, S., Ebrahimi, C. T., & Ruglass, L. M. (2022). Measurement error-corrected estimation of clinically significant change trajectories for interventions targeting comorbid PTSD and SUDs in OEF/OIF Veterans. *Behavior Therapy*, *53*(5), 1009–1023. https://doi.org/10.1016/j.beth.2022.04.007

Saavedra, L. M., Morgan-López, A. A., Hien, D. A., Back, S. E., Killeen, T. K., Ruglass, L. M., López-Castro, T., & Lopez-Castro, T. (2021). Putting the patient back in clinical significance: Using item response theory in estimating clinically significant change in treatment for PTSD and SUDs. *Journal of Traumatic Stress*, *34*(2), 454–466. https://doi.org/10.1002/jts.22624

Saavedra, L. M., Morgan-López, A. A., Hien, D. A., López-Castro, T., Ruglass, L. M., Back, S. E., Fitzpatrick, S., Norman, S. B., Killeen, T. K., Ebrahimi, C. T., Hamblen, J., & the Consortium on Addictions, Stress, and Trauma. (2021). Evaluating treatments for posttraumatic stress disorder, alcohol and other drug use disorders using meta-analysis of individual patient data: Design and methodology of a virtual clinical trial. *Contemporary Clinical Trials*, *107*, 106479. https://doi.org/10.1016/j.cct.2021.106479

Saladin, M. (unpublished). Trial comparing propranolol to placebo for treating PTSD and alcohol use disorders. Unpublished dataset.

Sannibale, C., Teesson, M., Creamer, M., Sitharthan, T., Bryant, R. A., Sutherland, K., Taylor, K., Bostock-Matusko, D., Visser, P., Peek-O'Leary, M. (2013). Randomized controlled trial of cognitive behaviour therapy for comorbid post-traumatic stress disorder and alcohol use disorders. *Addiction*, *108*, 1397–1410. https://doi.org/10.1111/add.12167

Schacht, R. L., Brooner, R. K., King, V. L., Kidorf, M. S., & Peirce, J. M. (2017). Incentivizing attendance to prolonged exposure for PTSD with opioid use disorder patients: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *85*, 689.

Schäfer, I., Lotzin, A., Hiller, P., Sehner, S., Driessen, M., Hillemacher, T., Schafer, M., Scherbaum, N., Schneider, B., & Grundmann, J. (2019). A multisite randomized controlled trial of Seeking Safety vs. Relapse Prevention Training for women with co-occurring posttraumatic stress disorder and substance use disorders. *European Journal of Psychotraumatology*, *10*(1), 1577092. https://doi.org/10.1080/20008198.2019.1577092

Silverstein, M. W., Petri, J. M., Kramer, L. B., & Weathers, F. W. (2020). An item response theory analysis of the PTSD checklist for DSM-5: Implications for DSM-5 and ICD-11. *Journal of Anxiety Disorders*, *70*, 102190. https://doi.org/10.1016/j.janxdis.2020.102190

Simpson, T. L., Rise, P., Browne, K. C., Lehavot, K., & Kaysen, D. (2019). Clinical presentations, social functioning, and treatment receipt among individuals with comorbid life-time PTSD and alcohol use disorders versus drug use disorders: Findings from NESARC-III. *Addiction*, *114*(6), 983–993. https://doi.org/10.1111/add.14565

Sinharay, S., & Haberman, S. J. (2014). How often is the misfit of item response theory models practically significant? *Educational Measurement: Issues and Practice*, *33*(1), 23–35. https://doi.org/10.1111/emip.12024

Sonne, S. (unpublished). Trial comparing Paxil to placebo for treating PTSD and substanceuse disorders. Unpublished dataset.

Spoont, M. R., Hodges, J., Murdoch, M., & Nugent, S. (2009). Race and ethnicity as factors in mental health service use among veterans with PTSD. *Journal of Traumatic Stress*, *22*(6), 648–653. https://doi.org/10.1002/jts.20470

Spoont, M., Nelson, D., Van Ryn, M., & Alegría, M. (2017). Racial and ethnic variation in perceptions of VA mental health providers are associated with treatment retention among veterans with PTSD. *Medical Care*, *55*(Suppl 2), S33–S42. https://doi.org/10.1097/mlr.0000000000000755

Street, A. E., Vogt, D., & Dutra, L. (2009). A new generation of women veterans: Stressors faced by women deployed to Iraq and Afghanistan. *Clinical Psychology Review*, *29*(8), 685–694. https://doi.org/10.1016/j.cpr.2009.08.007

Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1–67. https://doi.org/10.18637/jss.v045.i03

Vujanovic, A. A., Smith, L. J., Green, C. E., Lane, S. D., & Schmitz, J. M. (2018). Development of a novel, integrated cognitive-behavioral therapy for co-occurring posttraumatic stress and substance use disorders: A pilot randomized clinical trial. *Contemporary Clinical Trials*, *65*, 123–129. https://doi.org/10.1016/j.cct.2017.12.013

Waldrop, A. E., Back, S. E., Verduin, M. L., & Brady, K. T. (2007). Triggers for cocaine and alcohol use in the presence and absence of posttraumatic stress disorder. *Addictive Behaviors*, *32*(3), 634–639. https://doi.org/10.1016/j.addbeh.2006.06.001

Weathers, F. W., Bovin, M. J., Lee, D. J., Sloan, D. M., Schnurr, P. P., Kaloupek, D. G., Keane, T. M., & Marx, B. P. (2018). The Clinician-

Administered PTSD Scale for DSM–5 (CAPS-5): Development and initial psychometric evaluation in military veterans. *Psychological Assessment*, *30*(3), 383–395. https://doi.org/10.1037/pas0000486

Zlotnick, C., Johnson, J., & Najavits, L. M. (2009). Randomized controlled pilot study of cognitive-behavioral therapy in a sample of incarcerated women with substance use disorder and PTSD. *Behavior Therapy*, *40*(4), 325–336.

Zlotnick, C., Najavits, L. M., Rohsenow, D. J., & Johnson, D. M. (2003). A cognitive-behavioral treatment for incarcerated women with substance abuse disorder and posttraumatic stress disorder: Findings from a pilot study. *Journal of Substance Abuse Treatment*, *25*(2), 99–105.

How to cite this article: Morgan-López, A. A., Saavedra, L. M., Hien, D. A., Norman, S. B., Fitzpatrick, S. S., Ye, A., Killeen, T. K., Ruglass, L. M., Blakey, S. M., & Back, S. E. (2023). Differential symptom weighting in estimating empirical thresholds for underlying PTSD severity: Toward a "platinum" standard for diagnosis? *International Journal of Methods in Psychiatric Research*, *32*(3), e1963. https://doi.org/10.1002/mpr.1963