Routledge
Taylor & Francis Group

Journal of Personality Assessment

ISSN: 0022-3891 (Print) 1532-7752 (Online) Journal homepage: www.tandfonline.com/journals/hjpa20

Multivariate Base Rates of Standard- and Skyline-Cutoff Elevations on the Personality Assessment Inventory: Do They Distinguish Simulated from Genuine PTSD?

Stephen L. Aita, Emily L. Montgomery, Joshua E. Caron, Louis A. Pagano Jr., Michael J. Broggi, Paul B. Ingram, Steven C. Erickson, Nicholas C. Borgogna, Grant G. Moncrief, Robert M. Roth, Matthew R. Calamia, Patrick Armistead-Jehle & Benjamin D. Hill

To cite this article: Stephen L. Aita, Emily L. Montgomery, Joshua E. Caron, Louis A. Pagano Jr., Michael J. Broggi, Paul B. Ingram, Steven C. Erickson, Nicholas C. Borgogna, Grant G. Moncrief, Robert M. Roth, Matthew R. Calamia, Patrick Armistead-Jehle & Benjamin D. Hill (21 Jul 2025): Multivariate Base Rates of Standard- and Skyline-Cutoff Elevations on the Personality Assessment Inventory: Do They Distinguish Simulated from Genuine PTSD?, Journal of Personality Assessment, DOI: 10.1080/00223891.2025.2530047

To link to this article: https://doi.org/10.1080/00223891.2025.2530047

JOURNAL OF PERSONALITY ASSESSMENT
https://doi.org/10.1080/00223891.2025.2530047

Multivariate Base Rates of Standard- and Skyline-Cutoff Elevations on the Personality Assessment Inventory: Do They Distinguish Simulated from Genuine PTSD?

Stephen L. Aita[1,2], Emily L. Montgomery[2], Joshua E. Caron[2,3], Louis A. Pagano Jr.[4], Michael J. Broggi[2], Paul B. Ingram[5], Steven C. Erickson[6], Nicholas C. Borgogna[7], Grant G. Moncrief[8,9], Robert M. Roth[8,9], Matthew R. Calamia[10], Patrick Armistead-Jehle[11] and Benjamin D. Hill[12]

[1]Department of Neurology, Whiddon College of Medicine at the University of South Alabama; [2]Department of Mental Health, VA Maine Healthcare System; [3]Department of Psychology, University of Maine; [4]Office of the Chief of Staff, St. Cloud VA Healthcare System; [5]Department of Psychological Sciences, Texas Tech University; [6]Department of Mental Health, VA Durham Healthcare System; [7]Department of Psychology, University of Alabama at Birmingham; [8]Department of Psychiatry, Dartmouth Health; [9]Department of Psychiatry, Geisel School of Medicine at Dartmouth; [10]Department of Psychology, Louisiana State University; [11]Concussion Clinic, Munson Army Health Center; [12]Department of Psychology, University of South Alabama

ABSTRACT

Multivariate base rates (MBR) of elevations are an emerging psychometric paradigm for enhanced interpretation of multiscale self-report data. The aims of this study were to calculate and compare MBR of scale/subscale elevations on the Personality Assessment Inventory (PAI) and determine the ability of MBR to differentiate between mood disorders ($n = 524$, $k = 3$), military-based posttraumatic stress disorder (PTSD; $n = 252$, $k = 2$), and coached PTSD-simulator ($n = 160$, $k = 1$) groups. Overall, having *at least* one standard ($T \geq 70$) and skyline elevation on clinical scales and clinical subscales was common across the groups. However, differential abnormal elevation thresholds emerged for each group. For instance, it was *unusual* (i.e., MBR < 10%) for the mood disorders group to have ≥ 1 (9.7%) and for the genuine PTSD group to have ≥ 3 (9.1%) skyline-elevated clinical scales. For subscales, it was *unusual* for the mood and PTSD groups to have ≥ 3 (7.6%) and ≥ 7 (8.3%) skyline-elevated clinical subscales, respectively. Conversely, PTSD simulators commonly yielded profiles with standard- and skyline elevations on nearly all clinical scales and subscales. MBR cutoffs identified from receiver-operating characteristic curve analyses yielded robust sensitivity (.650-.806) and specificity (.833-.984) in differentiating genuine PTSD and mood disorder groups from PTSD simulators. MBR are useful in differentiating genuine from simulated psychopathology, consistent with broader scale-based infrequency approaches.

ARTICLE HISTORY
Received 5 February 2025
Revised 23 May 2025
Accepted 23 June 2025

In the U.S., posttraumatic stress disorder (PTSD has a lifetime prevalence of around 6% but higher (9.4%) in military personnel (Wisco et al., 2022). Approximately 22% of Veterans from the conflicts in Iraq and Afghanistan have experienced PTSD symptoms (Seal et al., 2009). In turn, PTSD is the most diagnosed mental health disorder among Veterans (Department of Veterans Affairs, 2015). The Annual Benefits Report from the US Department of Veterans Affairs (2023) indicated that PTSD accounts for nearly 60% of all mental health disorder compensation claims, resulting in 57,998 new PTSD compensation recipients in 2022 alone. With compensation benefits, however, comes large incentives to feign symptoms, and PTSD is a common claim in personal injury litigation and disability evaluations. Risk of non-credible claims is high in disability evaluation settings (Rogers & Bender, 2020), with estimates of feigned presentations around 30% in civil litigation cases juxtaposed with ~15% in clinical settings (Gervais et al., 2001; Green et al., 2001; Larrabee et al., 2009; Mittenberg et al., 2002; Proto et al., 2014; Young, 2015). Several studies report higher prevalence of symptom exaggeration for PTSD-related disability claims, with base rates ranging from 30 to 53% (Freeman et al., 2008; Lees-Haley, 1997; Taylor et al., 2007). In turn, developing robust methods for detecting feigned PTSD is essential to preserve treatment resources and discourage fraud by non-credible claimants.

Symptom validity within PTSD

The Personality Assessment Inventory (PAI; Morey, 2007) is a widely used standardized instrument for evaluating PTSD among other mental health concerns (Calhoun et al., 2010). It provides four categories of information: validity of responses, diagnostic/clinical indications, interpersonal styles, and treatment prognosis. Examinees with PTSD often yield profiles with elevations across several clinical scales such as

CONTACT Stephen L. Aita Slaita@health.southalabama.edu Department of Neurology, Whiddon College of Medicine at the University of South Alabama, 1601 Center St, Mobile, AL 36604, USA.
Supplemental data for this article can be accessed online at https://doi.org/10.1080/00223891.2025.2530047.

Somatic Complaints, Anxiety and Related Disorders, Anxiety, Depression, and Schizophrenia (Mozley et al., 2005; Rogers et al., 1996). Numerous PAI embedded symptom validity measures have been created over time and studied in the setting of feigned PTSD, such as the Negative Impression Management (NIM) scale, Malingering Index (MAL), Rogers Discriminant Function (RDF), Negative Distortion Scale (NDS), Hong Malingering Function (HMF), and Multiscale Feigning Index (MFI; Hong & Kim, 2001; Mogge et al., 2010; Morey, 2007; Rogers et al., 1996).

Large between-group (i.e., true vs. feigned PTSD) effects (e.g., *d*'s exceeding 1.00) are generally seen for the early PAI negative distortion indicators NIM and MAL (Calhoun et al., 2010; Liljequist et al., 1998; Scragg et al., 2000), whereas newer scales (such as HMF and MFI) appear to yield even stronger findings (Rogers et al., 1993; Russell & Morey, 2019; Thomas et al., 2012). For instance, Thomas et al. (2012) found large group differences between PTSD feigners and genuine PTSD groups on NIM ($d = 1.24$), MAL ($d = 1.28$), RDF ($d = 0.82$), and NDS ($d = 1.60$). These negative distortion indicators correctly classified 74-97% of genuine PTSD cases and 26-78% of PTSD feigners, with NDS (at a cutoff of $T > 85$) performing the best (correctly classifying 97% of genuine PTSD and 64% of feigners). Later research also observed robust mean differences between feigned and genuine PTSD groups on NIM ($d = 0.68$), MAL ($d = 0.73$), RDF ($d = 0.60$), NDS ($d = 0.66$), HMF ($d = 0.95$), and MFI ($d = 1.39$; Russell & Morey, 2019). However, sensitivity and specificity varied considerably across indexes and was best for the MFI (59.1% and 92.0%, respectively).

In a separate vein, research on false positive malingering classifications suggest between 13 and 26% of PAI profiles classify "true" PTSD patients as invalid when using traditional validity scales (Calhoun et al., 2000; Mozley et al., 2005). According to Calhoun et al. (2000), the recommended cutoffs for the MAL and NIM produced 26.6% and 57.8% false positives classifications for PTSD cases, respectively. Mozley et al. (2005) found 27.2% of their genuine PTSD sample was misclassified as non-credible using the RDF's recommended cutoff. Thus, developing alternative methods that increase sensitivity while minimizing false positive classifications is warranted.

One promising approach for detecting feigned PTSD is examining the rate of multiple elevations across scales and subscales, commonly known as multivariate base rates (MBR). Logistic regression or discriminant function equations for predicting symptom exaggeration often prove difficult to replicate but MBR are generally stable in populations (Aita et al., 2023; Brooks et al., 2013; Crawford et al., 2011; Oltra-Cucarella et al., 2021). MBR approaches also minimize misclassification rates associated with methods relying on single scale elevations. Considering the variability often observed in "true" PTSD PAI profiles, including the tendency for those with PTSD to produce elevations on multiple clinical scales, it may be possible to detect differences between legitimate and non-credible responding by considering both the number and severity of elevations within the PAI profile using an MBR approach. While no symptom validity research has applied MBR to the PAI, Gaines et al. (2013) was the first on the PAI to consider elevations across multiple scales as an embedded symptom validity indicator. They averaged *T*-scores from seven PAI clinical scales to develop the MFI, reported its promising sensitivity (68.89%) and specificity (94.34%) for detecting feigning in their initial study, and found it performed better than existing indicators such as MAL, NIM, and RDF (Gaines et al., 2013).

Primer on base rates and multivariate base rates

Base rates are obtained by dividing the number of cases with a clinical sign of interest (e.g., an elevated score on a depression screening measure) by the number of cases being analyzed (Aita et al., 2018), then multiplying the resultant quotient by 100 (Aita et al., 2022). Thus, if five of one-hundred cases yield an elevated depression screening score, then the base rate of the positive finding for this specified group of cases is 5%. Base rates inform decision making and diagnostic accuracy of instruments in a variety of clinical psychologic contexts (Labarge et al., 2003; Meehl & Rosen, 1955; Palmer et al., 1998). Indeed, clinical conditions with low and high base rates differentially impact how diagnostic tests function, especially their predictive value (Eisenberg, 1995).

Authors caution against applying univariate base rates (UBR) to scores within test batteries or multiscale instruments as these usually underestimate the likelihood of obtaining an abnormal score (Aita et al., 2022; Brooks et al., 2013; Karr et al., 2017, 2018). To demonstrate, according to the Gaussian distribution, 5% of typical examinees can be expected to produce one *low* score (i.e., $j = 1$, with a low score operationalized as scaled score ≤ 5 or *z*-score ≤ -1.7) on the Wechsler Memory Scale-Fourth Edition (Wechsler, 2009). However, administration of this measure yields eight primary memory scores (i.e., $k = 8$), and when simultaneously interpreting all such scales, the likelihood of an examinee producing *at least* one low score ($j \geq 1$) is about 29% (juxtaposed to 5% if examining one index score *in isolation*; Brooks et al., 2013). Conceptually, the disparity between probabilities in this illustration speaks to a fundamental problem of applying a singular, univariate Gaussian probability to a clinical model in which multiple test scores are simultaneously administered and interpreted.

Whereas UBR is a good fit for interpreting single indicators in isolation, MBR provides a solution for simultaneously interpreting probabilistic information from multiple indicators (e.g., a psychodiagnostic measure with numerous scales or determining cumulative exposure across multiple trauma types; Borgogna et al., 2023; Karr et al., 2017, 2018). MBR is conceptualized as "…the likelihood of obtaining low scores, when several scores are administered and interpreted…" (Brooks et al., 2013; p. 76). MBR have traditionally been applied to standardization samples of various performance-based cognitive measures such as the Wechsler Adult Intelligence Scale and Wechsler Memory Scale (Brooks et al., 2011). Collectively, these studies demonstrate that it is common (e.g., 20-40% when defining a low score as $z \leq -1.7$, allowing for variation of how many scores are interpreted) for healthy examinees to produce *some* low scores when considering multiple cognitive test scores.

Few studies have applied MBR to multidimensional rating scales. Aita et al. (2023) examined MBR of elevated Behavior Rating Inventory of Executive Function-Adult Version (BRIEF-A; Roth et al., 2005) scores in the standardization samples ($n = 2{,}250$). They found that UBR ranged from 3 to 16% when interpreting primary scales in isolation. Conversely, when simultaneously considering the nine primary BRIEF-A scales, MBR of obtaining *at least* one ($j \geq 1$) elevated score were much higher, ranging from 23 to 37%. This same group largely replicated these findings using the standardization samples ($n = 3{,}603$) for the BRIEF, Second Edition (BRIEF2; Gioia et al., 2015), where UBR on scales ranged from 7 to 12% and MBR ranged from 26 to 27% (Aita et al., 2022). Recently, Ingram and Karr (2024) examined MBR of "potentially problematic" scores (defined as ±1 *SD* from the mean, "depending on the orientation of the scale") on the NIH Toolbox Emotion Battery ($k = 17$ scales, $k = 3$ subdomains) in the normative sample ($n = 753$). They found a nontrivial probability of normative respondents obtaining *at least* one elevated scale (61.2%) and subdomain (23.2%) on this self-report measure. Together, simple base rate information consistently underestimates the likelihood of obtaining abnormal scores when numerous scores from multiscale rating measures are considered.

Critically, only two studies have applied MBR to rating instruments using multi-clinical group data. Shura et al. (2022) presented MBR of obtaining elevated ($T \geq 65$) scores across the RC scales ($k = 8$) on the Minnesota Multiphasic Personality Inventory-2-Restructured Form for a sample of Veterans who were assessed at a VA Attention-Deficit/Hyperactive Disorder (ADHD) specialty clinic. Obtaining *at least* one ($j \geq 1$) elevated RC scale score was common in the total sample (MBR = 75.3%), and MBR were similar in their subsamples: healthy = 62.1%, no ADHD = 80.6, any ADHD = 71.5%, ADHD-Inattentive = 67.4%, and ADHD-Combined = 82.9%. However, as number of elevated scores (j) increased, MBR were higher in their clinical groups. To illustrate, the MBR of obtaining four *or more* ($j \geq 4$) elevated RC scale scores was over twice as high in the no ADHD (28.0%) and any ADHD (21.5%) groups relative to healthy Veterans (10.3%; Shura et al., 2022). Next, Aita et al. (2022) studied MBR of elevated scores on the BRIEF2 using clinical standardization data from children with Specific Learning Disorder in Reading (SLD-R), ADHD, and autism spectrum disorder (ASD). They found these clinical groups had anywhere between a two- to four-fold increase in MBR of having *at least* one ($j \geq 1$) elevated score relative to the non-clinical standardization samples, and that MBR increased with clinical severity such that those with ADHD (52-84%) and ASD (55-83%) had higher MBR than SLD-R (27-44%; Aita et al., 2022).

The present study

MBR approaches enhance interpretation of rating scale measures, particularly likelihoods of obtaining abnormal scores when multiple scores are considered. However, this emerging psychometric paradigm has not been applied to the PAI. Moreover, no studies have explored MBR as a tool to help classify non-credible responding. The PAI is ideally positioned for application of MBR given its multidimensional nature, embedded symptom validity indicators, and documented clinical efficacy for use in PTSD. The aims of the current study were therefore two-fold: (1) describe MBR of elevated scores across participants with mood disorders, PTSD, and simulated PTSD; (2) determine whether MBR of elevated PAI scores can identify non-credible PAI profiles among PTSD simulators while differentiating them from those with genuine PTSD. As a supplemental aim, we also compared diagnostic effectiveness of MBR to the NIM and MFI negative distortion scales. Broadly, this study sought to provide clinicians with an approach for simultaneously interpreting multiple scores on the PAI, as well as an easy-to-apply method for identifying non-credible PAI profiles in the context of PTSD evaluations. We hypothesized there would be a graded pattern of increasing MBR (mood disorder < PTSD < PTSD simulators) and that MBR of profile elevations would yield acceptable diagnostic accuracy (sensitivity ≥ .50 and specificity ≥ .90; Larrabee et al., 2019) in identifying non-credible PTSD.

Method

Participants

This cross-sectional study featured data from Veterans, active-duty Service Members (SM), and civilians, all of which were presenting for clinical evaluations. In addition, a sample of college students were coached to simulate PTSD. Data were ultimately condensed into the following groups: mood disorders, PTSD, and PTSD simulators. Diagnoses for all cases in the mood disorders and PTSD groups were established using established DSM-based criteria (American Psychiatric Association, 2000, 2013). All data were collected with approval of the relevant Institutional Review Boards.

The PTSD group was comprised of Veterans ($n = 111$) and SMs ($n = 141$). Data from the Veterans were identified through retrospective chart reviews. Veterans were included if they had a diagnosis of PTSD either given at the time the clinical evaluation or if they had a PTSD diagnosis under Active Problems on CPRS in the VA system. Of the included Veterans, about one-quarter ($n = 40$, 28.0%) were undergoing compensation and pension evaluations; the rest were administered the PAI as part of routine clinical evaluations. Most ($n = 110$, 76.9%) were 50-100% service connected. SM data were drawn from a larger database of SMs who presented for routine clinical neuropsychological evaluations at a Midwest Army Health Center. For the PTSD group, cases were not limited to PTSD-monodiagnosis, though cases with schizophrenia and major neurocognitive disorders (i.e., dementia) were excluded.

Next, we elected to include a mood disorders group to enhance the contrasts of MBR against examinees with genuine and simulated PTSD. This group was comprised of Veterans ($n = 32$), SMs ($n = 361$), and civilian adults ($n = 131$). These individuals were diagnosed with depressive, anxiety-related, and/or adjustment disorders and

PTSD diagnoses were excluded (see Supplemental Tables 1–3 for breakdown of psychiatric comorbidities). The civilian mood disorders sample came from a university-based community outpatient training clinic where they had undergone comprehensive psychological evaluations for a variety of referral reasons. For all patients in the mood disorder group, we only considered cases with mood disorders without comorbid serous mental illness (e.g., bipolar disorder, schizophrenia spectrum disorders), substance use disorders, or neurocognitive and neurodevelopmental disorders.

PTSD simulators were drawn from a sample of college students who received research credit for participation. These were coached to simulate PTSD using a vignette adapted from prior research (Hill et al., 2015; Musso et al., 2016). See Supplemental Table 4 for the vignette. Prior to creating the merged diagnostic groups used in this project, Mann-Whitney *U* tests were used to compare number of standard and skyline elevations on PAI clinical scales and subscales (which were used to compute MBR indicators as described below) among subsamples: PTSD (i.e., PTSD+ Veterans vs. PTSD+ SMs) and mood disorders (i.e., mood disorder + Veterans vs. mood disorder + SMs vs. mood disorder + civilians). Number of PAI scale and subscale elevations did not meaningfully differ (according to the Ferguson, 2009 criteria described below) among subsamples compared respective diagnostic groups. As such, PTSD and mood disorders subgroups from different samples were combined as planned. All PAIs were administered by or under supervision of licensed psychologists.

Measures

All participants completed the PAI as a part of a neuropsychological or psychological assessment. The PAI is a broad-band self-report measure of personality and psychopathology that was developed with a focus on maximizing content validity consistent with the DSM-IV (Morey, 2007). It is composed of 344 items, each with a four-point response scale (*False, Slightly True, Mostly True, Very True*). The PAI items are organized into four response-validity scales, 11 clinical syndrome scales (nine of which have subscales), five treatment indicator scales, and two interpersonal scales.

Protocols analyzed in this project were electronically scored. Raw scores were converted to *T*-scores ($M = 50$, $SD = 10$) based on data from the original census-matched (for age, sex, and race) normative PAI sample of $n = 1{,}000$ community-dwelling adults. Higher values on the PAI generally reflect adverse outcomes (e.g., a greater endorsed level of psychopathology symptoms). The current study used the "standard" cutoff of $T \geq 70$ for a clinical elevation. In addition, we examined a second elevation cutoff for each PAI indicator based on the "clinical skyline" which reflects scores falling two standard deviations above the clinical PAI sample ($n = 1{,}256$ clinical patients mixed from inpatient and outpatient psychiatric settings). A comprehensive list of scale and subscale names, abbreviations, and skyline elevation cutoffs are presented on Supplemental Tables 5 and 6.

Analysis plan

Prior to proceeding with computation of MBR, cases were removed if PAI protocols were not interpretable according to elevated inconsistency (ICN ≥ 73) and/or unusual/bizarre statements endorsed (INF ≥ 75; Morey, 2007). The following cases were removed for such reasons: mood disorder ($n = 20$, 3.7%), PTSD ($n = 10$, 3.8%), and the PTSD-simulator ($n = 55$, 25.6%) groups. In addition, group demographic factors were compared using chi-squared tests of independence for categorical variables and one-way ANOVAs (with Tukey's HSD post-hoc pairwise comparisons for significant models) for continuous variables. Next, bivariate correlations between demographic variables and MBR outcomes (specifically, number of standard and skyline elevations [*j*] across scales and subscales) were tested using Spearman's rho (ρ). *Practically* significant demographic-group and demographic-MBR relations were interpreted using Ferguson's (2009) guidelines: $\varphi_c \geq .20$, $\eta^2 \geq .04$, $d \geq 0.41$, $r \geq .20$, and $\rho \geq .20$.

We followed previously published guidelines for computing MBR on self-report measures (Aita et al., 2022; 2023). First, we examined MBR of elevated scores across the PAI clinical scales: SOM, ANX, ARD, DEP, MAN, PAR, SCZ, BOR, ANT, ALC, and DRG ($k = 11$ scales). UBR (i.e., proportion of cases with an elevated score on a single indicator) of elevated *T*-scores were computed for each specific scale. Next, the number of elevated scores were tallied across scales to derive "*j*" (i.e., the number of abnormal scores). Finally, to obtain *cumulative* MBR of elevated scores, probabilities were calculated for cases with one *or more* elevated scores (i.e., $j \geq 1$), two *or more* elevated scores (i.e., $j \geq 2$), and so on until a final non-cumulative MBR was calculated for cases with elevated scores on all scales (i.e., $j = 11$). Probabilities were compounded according to the specified number of elevations. For instance, the MBR of eight *or more* elevated scores (i.e., $j \geq 8$) is obtained by: $P(j = 8) + P(j = 9) + P(j = 10) + P(j = 11)$, where $P(j = 11)$ denotes the prevalence of cases with 11 elevations relative to the entire sample (i.e., $n_{j=11}/N$). This same approach was repeated on the PAI clinical subscales to obtain MBR of elevations on these: SOM-C, SOM-S, SOM-H, ANX-C, ANX-A, ANX-P, ARD-O, ARD-P, ARD-T, DEP-C, DEP-A, DEP-P, MAN-A, MAN-G, MAN-I, PAR-H, PAR-P, PAR-R, SCZ-P, SCZ-S, SCZ-T, BOR-A, BOR-I, BOR-N, BOR-S, ANT-A, ANT-E, ANT-S ($k = 28$ subscales). Separate MBR were computed for standard and skyline elevations. We followed Brooks et al. (2011) proposed convention in defining MBR of < 10% as reflecting an *unusual* (or *well below expected*) finding. To illustrate, if less than 10% of examinees in a specified group obtain three or more ($j \geq 3$) standard elevations on PAI clinical scales, we would label this MBR as *unusual*.

Next, we performed a series of receiver operating characteristic (ROC) analyses to test the diagnostic efficacy of the various MBR metrics in discerning between genuine and simulated (i.e., non-credible) PTSD presentations based on standard and skyline elevations on PAI profiles. Specifically, separate area under the ROC curve (AUC) values were computed for the following indicators: number of (i.e., *j*) standard

scale elevations, skyline scale elevations, standard subscale elevations, and skyline subscale elevations. The PTSD group was coded as the referent, and PTSD simulators coded as the target group. We also conducted an additional set of ROC analyses on the same MBR variables to see if these showed utility in differentiating participants in the mood disorders group (coded as referent) from PTSD simulators (coded as the target group). Adjunctive to MBR variables, the NIM PAI validity scale and MFI were analyzed to compare their relative classification effectiveness (in discerning genuine PTSD or mood disorders from simulated PTSD) to MBRs. ROC model diagnostic quality was assessed according to the strength of AUC coefficients: < .70 = "poor," .70 to .90 = "moderate," and > .90 = "high" classification accuracy (Swets, 1988).

We then identified optimal cutoffs of number of elevations (i.e., *j*) for MBR variables using the Youden's index, which reflects the point at which sensitivity and specificity are maximized (Youden, 1950). For each cutoff identified in ROC models, adjunctive to sensitivity and specificity, we computed Youden's *J* statistic, which ranges from 0 to 1 with higher values indicating greater diagnostic performance of the cutoff (Youden, 1950). We also obtained overall diagnostic accuracy by taking the ratio of correctly identified cases (true positives + true negatives) from all cases modeled (Baratloo et al., 2015). Next, we evaluated diagnostic efficiency statistics for the recommended cutoffs of NIM $T \geq 92$ (Morey, 2007) and MFI $T \geq 77$ (Gaines et al., 2013) for detecting feigning. Finally, we examined the degree of agreement among MBR, NIM, and MFI cutoffs using Cohen's kappa statistic (κ), and nested analyses within the ROC models (i.e., PTSD vs. PTSD simulators, mood disorders vs. PTSD simulators). We interpreted κ values as 0-0.20 = slight, 0.21-0.40 = fair, 0.41-0.60 = moderate, 0.61-0.80 = substantial, and 0.80-1.00 = almost perfect agreement (Landis & Koch, 1977).

Results

Final sample characteristics

Table 1 displays demographic characteristics for the primary three groups (i.e., mood disorders, PTSD, and PTSD simulators). *Practically* significant (Ferguson, 2009) group differences were observed for age and sex, where the simulator group was younger and had a higher proportion of female participants than the mood disorders and PTSD groups. These demographic factors (as well as race) were unrelated to MBR, NIM, and MFI outcomes in the mood disorders and PTSD groups (ρ's < .18). Age was weakly inversely related to MBR, NIM, and MFI outcomes in the PTSD-Simulator group (ρ's = −0.11 to −0.25). In the whole study sample, MBR variables were correlated with NIM (ρ's = .67 to .83), and to a greater extent, MFI (ρ's = .72 to .96; see Supplemental Table 7). Within the SM study population, history of concussion was not associated with number of scale and subscale elevations (ρ's = −0.02 to −0.04).

Univariate base rates on clinical scales

Table 2 presents UBR of standard and skyline elevations for the PAI scales. In addition, Supplemental Tables 8 ad 9 present descriptive statistical information (*M*, *SD*) for PAI clinical scales and clinical subscales, respectively, stratified by group. In the mood disorders group, the highest BRs of standard elevation were seen on DEP (33.2%), ANX (29.0%), SCZ (26.0%) and ARD (21.2%). Conversely, skyline elevations across clinical scales were infrequent for the mood disorders group, with all UBRs falling below 6.0%. Next, for the PTSD group, highest UBRs of standard elevations were seen on the DEP (65.5%), ARD (62.3%), ANX (50.8%), SOM (44.8%), and SCZ (44.8%) scales. Aside from the PAR (12.7%), SOM (12.3%), and ARD (11.1%) scales, UBRs of skyline elevations were unusual, all falling below 10%. Lastly, among PTSD simulators, relatively larger UBRs of standard scale elevations were observed for most clinical scales, with highest rates on ANX (88.8%), SUI (81.3%), SCZ (88.1%), and DEP (87.5%). Similarly, skyline scale elevations were common among PTSD simulators with highest UBRs on SCZ (61.3%), PAR (53.1%), and ARD (48.1%). Of note, UBR of elevated NIM was rare for the mood disorder (1.3%) and PTSD (5.6%) groups but encompassed over half the simulators (65.0%). Elevated MFI was only rare in the mood disorder group (3.2%) but not PTSD (21.0%) and simulators (76.3%).

Table 1. Sample demographic characteristics.

	Group			Omnibus comparison	
Variable	Mood dx *n* = 524	PTSD *n* = 252	PTSD sim *n* = 160	*p*	*ES*
Age, *M* (*SD*)	34.0 (10.7)	41.0 (10.6)	20.0 (3.9)	<.001	$\eta^2 = .325$
Sex, *n* (%)					
Male	381 (72.7)	219 (86.9)	42 (26.3)	<.001	$\varphi_c = .434$
Female	143 (27.3)	33 (13.1)	118 (73.8)		
Race/ethnicity, *n* (%)					
Black	66 (12.6)	36 (14.3)	15 (9.4)	.007	$\varphi_c = .121$
White	413 (78.8)	194 (77.0)	128 (80.0)		
Hispanic	26 (5.0)	13 (5.2)	4 (2.5)		
Native American	1 (0.2)	3 (1.2)	0 (0.0)		
Asian	11 (2.1)	4 (1.6)	6 (3.8)		
Pacific Islander	5 (1.0)	1 (0.4)	1 (0.6)		
Multiracial	2 (0.4)	1 (0.4)	6 (3.8)		
Education (years), *M* (*SD*)	15.0 (2.4)	14.4 (2.4)	13.6 (0.9)	<.001	$\eta^2 = .053$

Note. PTSD: Posttraumatic Stress Disorder; PTSD sim: PTSD Simulators; *ES*: Effect Size. *p*- and *ES*-values reflect one-way ANOVAs for continuous variables (age, education), and chi-squared tests of independence for categorical variables (sex, race/ethnicity).

Table 2. Univariate base rates of standard (left) and skyline (right) cutoff elevations on the PAI clinical scales.

Standard elevations				Skyline elevations			
	Group, *n* (%)				Group, *n* (%)		
PAI scale	Mood dx *n* = 524	PTSD *n* = 252	PTSD sim *n* = 160	PAI scale	Mood dx *n* = 524	PTSD *n* = 252	PTSD sim *n* = 160
				Validity:			
				NIM *T* ≥ 92	7 (1.3)	14 (5.6)	104 (65.0)
				PIM *T* ≥ 68	12 (2.3)	2 (0.8)	0 (0.0)
				MFI *T* ≥ 77	17 (3.2)	53 (21.0)	122 (76.3)
Clinical:				Clinical:			
SOM *T* ≥ 70	84 (16.0)	113 (44.8)	131 (81.9)	SOM *T* ≥ 88	6 (1.1)	31 (12.3)	69 (43.1)
ANX *T* ≥ 70	152 (29.0)	128 (50.8)	142 (88.8)	ANX *T* ≥ 91	7 (1.3)	16 (6.3)	67 (41.9)
ARD *T* ≥ 70	111 (21.2)	157 (62.3)	137 (85.6)	ARD *T* ≥ 91	1 (0.2)	28 (11.1)	77 (48.1)
DEP *T* ≥ 70	174 (33.2)	165 (65.5)	140 (87.5)	DEP *T* ≥ 96	7 (1.3)	24 (9.5)	66 (41.3)
MAN *T* ≥ 70	50 (9.5)	23 (9.1)	37 (23.1)	MAN *T* ≥ 75	19 (3.6)	11 (4.4)	15 (9.4)
PAR *T* ≥ 70	81 (15.5)	85 (33.7)	134 (83.8)	PAR *T* ≥ 84	15 (2.9)	32 (12.7)	85 (53.1)
SCZ *T* ≥ 70	136 (26.0)	113 (44.8)	141 (88.1)	SCZ *T* ≥ 90	10 (1.9)	19 (7.5)	98 (61.3)
BOR *T* ≥ 70	82 (15.6)	82 (32.5)	138 (86.3)	BOR *T* ≥ 92	0 (0.0)	4 (1.6)	34 (21.3)
ANT *T* ≥ 70	35 (6.7)	25 (9.9)	109 (68.1)	ANT *T* ≥ 82	6 (1.1)	7 (2.8)	66 (41.3)
ALC *T* ≥ 70	33 (6.3)	34 (13.5)	122 (76.3)	ALC *T* ≥ 98	1 (0.2)	2 (0.8)	27 (16.9)
DRG *T* ≥ 70	6 (1.1)	18 (7.1)	121 (75.6)	DRG *T* ≥ 96	0 (0.0)	1 (0.4)	55 (34.4)
TX and IP:				TX and IP:			
AGG *T* ≥ 70	54 (10.3)	60 (23.8)	120 (75.0)	AGG *T* ≥ 83	11 (2.1)	20 (7.9)	75 (46.9)
SUI *T* ≥ 70	48 (9.2)	47 (18.7)	130 (81.3)	SUI *T* ≥ 101	5 (1.0)	9 (3.6)	68 (42.5)
STR *T* ≥ 70	40 (7.6)	30 (11.9)	99 (61.9)	STR *T* ≥ 91	1 (0.2)	0 (0.0)	12 (7.5)
NON *T* ≥ 70	72 (13.7)	60 (23.8)	121 (75.6)	NON *T* ≥ 88	9 (1.7)	6 (2.4)	47 (29.4)
RXR *T* ≥ 70	2 (0.4)	0 (0.0)	1 (0.6)	RXR *T* ≥ 63	30 (5.7)	10 (4.0)	3 (1.9)
DOM *T* ≥ 70	22 (4.2)	17 (6.7)	4 (2.5)	DOM *T* ≥ 72	13 (2.5)	11 (4.4)	2 (1.3)
WRM *T* ≥ 70	5 (1.0)	0 (0.0)	0 (0.0)	WRM *T* ≥ 70	5 (1.0)	0 (0.0)	0 (0.0)

Note. PAI: Personality Assessment Inventory; PTSD: Posttraumatic stress disorder; PTSD sim: PTSD simulators; MFI: Multiscale Feigning Index; TX and IP: Treatment Consideration and Interpersonal scales. See Supplemental Table 5 for scale acronym definitions.

Univariate base rates on clinical subscales

Table 3 presents UBR of standard and skyline elevations for the PAI clinical subscales. For the mood disorders group, the highest standard UBRs on subscales were SCZ-T (51.2%), ANX-C (34.7%), ANX-A (30.5%), and DEP-A (28.2%); having a skyline-elevated subscale was unusual in this group as all UBRs fell below 10%. For the PTSD group, the highest standard UBRs on subscales were ARD-T (79.2%), SCZ-T (64.1%), ANX-A (56.8%), and DEP-A (56.8%). PTSD group UBRs of skyline elevations all fell below 10% except for PAR-H (15.1%), SCZ-S (14.1%), SOM-S (11.5%), ANX-P (10.9%), and PAR-R (10.4%). For simulators, most UBRs of standard-elevated subscales fell above 50% with the highest rates observed on ARD-T (89.4%), DEP-C (87.5%), ANX-A (86.9%), and ANX-P (86.3%). PTSD simulator respondents also commonly yielded skyline-elevated subscales, with highest UBRs observed on ARD-P (62.5%), ANX-P (58.1%), SCZ-P (53.8%), and PAR-P (51.2%).

Multivariate base rates on clinical scales

Table 4 presents cumulative MBR of standard and skyline elevations for the PAI clinical scales. Obtaining one *or more* ($j \geq 1$) standard elevation across scales was quite common, irrespective of group (mood disorders = 62.2%, PTSD = 85.3%, PTSD simulators = 92.5%). Next, it was unusual for examinees in the mood disorders group (7.3%) and PTSD group (7.9%) to obtain six or more ($j \geq 6$) and eight or more ($j \geq 8$) standard elevations on scales, respectively. Conversely, PTSD simulators *rarely* generated PAI profiles with *no* ($j = 0$) elevated scales (7.5%); that is, obtaining standard elevations across all scales ($j = 11$; 19.4%) was *not* uncommon in this group.

Having one *or more* ($j \geq 1$) skyline elevations on the PAI clinical scales was also common for the PTSD (32.5%) and PTSD simulator (77.5%) groups but not mood disorders (9.7%). Indeed, most (90.3%) examinees in the mood disorders group did not produce any ($j = 0$) skyline elevations across scales. Next, it was unusual for the PTSD and PTSD simulator groups to have three or more ($j \geq 3$; 9.1%) and 10 or more ($j \geq 10$; 8.8%) skyline elevations on scales, respectively.

Multivariate base rates on clinical subscales

Cumulative MBR of standard- and skyline-elevated scores on PAI clinical subscales ($k = 28$) are displayed on Table 5. As with the clinical scales, examinees across groups commonly obtained *at least* one standard-elevated score across subscales: mood disorders = 85.4%, PTSD = 94.8%, PTSD simulators = 94.4%. Thresholds for an *unusual* number of standard-elevated scores (j) across subscales were as follows: mood disorders $j \geq 12$ (8.9%), PTSD $j \geq 19$ (7.8%), and PTSD simulators $j \geq 27$ (7.5%).

With respect to MBR of skyline-elevated scores on PAI clinical subscales, examinees commonly had *at least* one skyline-elevated score in their response profile: mood disorders = 34.2%, PTSD = 50.5%, PTSD simulators = 86.9%. However, it was unusual for participants in the mood disorders and PTSD group to obtain three or more ($j \geq 3$; 7.6%) and seven or more ($j \geq 7$; 8.3%) of skyline-elevated scores, respectively. In stark contrast, it was *unusual* for those in the PTSD simulator group to obtain 20 or more ($j \geq 20$; 8.8%) skyline-elevated scores.

Table 3. Univariate base rates of standard (left) and skyline (right) cutoff elevations on the PAI clinical subscales.

Standard elevations				Skyline elevations			
	Group, *n* (%)				Group, *n* (%)		
PAI subscale	Mood dx *n* = 383	PTSD *n* = 192	PTSD sim *n* = 160	PAI subscale	Mood dx *n* = 383	PTSD *n* = 192	PTSD sim *n* = 160
SOM-C *T* ≥ 70	54 (14.1)	71 (37.0)	122 (76.3)	SOM-C *T* ≥ 93	8 (2.1)	19 (9.9)	68 (42.5)
SOM-S *T* ≥ 70	68 (17.8)	91 (47.4)	120 (75.0)	SOM-S *T* ≥ 86	10 (2.6)	22 (11.5)	53 (33.1)
SOM-H *T* ≥ 70	41 (10.7)	62 (32.3)	112 (70.0)	SOM-H *T* ≥ 85	7 (1.8)	14 (7.3)	42 (26.3)
ANX-C *T* ≥ 70	133 (34.7)	93 (48.4)	128 (80.0)	ANX-C *T* ≥ 89	5 (1.3)	8 (4.2)	31 (19.4)
ANX-A *T* ≥ 70	117 (30.5)	109 (56.8)	139 (86.9)	ANX-A *T* ≥ 91	6 (1.6)	12 (6.3)	47 (29.4)
ANX-P *T* ≥ 70	93 (24.3)	77 (40.1)	138 (86.3)	ANX-P *T* ≥ 89	13 (3.4)	21 (10.9)	93 (58.1)
ARD-O *T* ≥ 70	51 (13.3)	52 (27.1)	56 (35.0)	ARD-O *T* ≥ 78	15 (3.9)	8 (4.2)	24 (15.0)
ARD-P *T* ≥ 70	37 (9.7)	46 (24.0)	123 (76.9)	ARD-P *T* ≥ 79	8 (2.1)	12 (6.3)	100 (62.5)
ARD-T *T* ≥ 70	104 (27.2)	152 (79.2)	143 (89.4)	ARD-T *T* ≥ 99	1 (0.3)	17 (8.9)	48 (30.0)
DEP-C *T* ≥ 70	93 (24.3)	85 (44.3)	140 (87.5)	DEP-C *T* ≥ 96	7 (1.8)	17 (8.9)	68 (42.5)
DEP-A *T* ≥ 70	108 (28.2)	109 (56.8)	136 (85.0)	DEP-A *T* ≥ 99	6 (1.6)	15 (7.8)	43 (26.9)
DEP-P *T* ≥ 70	97 (25.3)	103 (53.6)	133 (83.1)	DEP-P *T* ≥ 86	5 (1.3)	15 (7.8)	44 (27.5)
MAN-A *T* ≥ 70	47 (12.3)	27 (14.1)	50 (31.3)	MAN-A *T* ≥ 76	24 (6.3)	10 (5.2)	22 (13.8)
MAN-G *T* ≥ 70	22 (5.7)	6 (3.1)	4 (2.5)	MAN-G *T* ≥ 74	10 (2.6)	4 (2.1)	2 (1.3)
MAN-I *T* ≥ 70	50 (13.1)	46 (24.0)	76 (47.5)	MAN-I *T* ≥ 78	20 (5.2)	19 (9.9)	44 (27.5)
PAR-H *T* ≥ 70	73 (19.1)	77 (40.1)	121 (75.6)	PAR-H *T* ≥ 83	14 (3.7)	29 (15.1)	66 (41.3)
PAR-P *T* ≥ 70	30 (7.8)	38 (19.8)	120 (75.0)	PAR-P *T* ≥ 83	10 (2.6)	16 (8.3)	82 (51.2)
PAR-R *T* ≥ 70	41 (10.7)	47 (24.5)	117 (73.1)	PAR-R *T* ≥ 81	15 (3.9)	20 (10.4)	76 (47.5)
SCZ-P *T* ≥ 70	25 (6.5)	17 (8.9)	103 (64.4)	SCZ-P *T* ≥ 80	9 (2.3)	7 (3.6)	86 (53.8)
SCZ-S *T* ≥ 70	85 (22.2)	88 (45.8)	129 (80.6)	SCZ-S *T* ≥ 87	15 (3.9)	27 (14.1)	72 (45.0)
SCZ-T *T* ≥ 70	196 (51.2)	123 (64.1)	133 (83.1)	SCZ-T *T* ≥ 93	22 (5.7)	12 (6.3)	55 (34.4)
BOR-A *T* ≥ 70	63 (16.4)	73 (38.0)	127 (79.4)	BOR-A *T* ≥ 91	3 (0.8)	3 (1.6)	25 (15.6)
BOR-I *T* ≥ 70	83 (21.7)	66 (34.4)	101 (63.1)	BOR-I *T* ≥ 89	3 (0.8)	0 (0.0)	11 (6.9)
BOR-N *T* ≥ 70	66 (17.2)	60 (31.3)	110 (68.8)	BOR-N *T* ≥ 87	9 (2.3)	6 (3.1)	21 (13.1)
BOR-S *T* ≥ 70	28 (7.3)	24 (12.5)	109 (68.1)	BOR-S *T* ≥ 88	3 (0.8)	5 (2.6)	43 (26.9)
ANT-A *T* ≥ 70	21 (5.5)	29 (15.1)	82 (51.2)	ANT-A *T* ≥ 84	2 (0.5)	1 (0.5)	19 (11.9)
ANT-E *T* ≥ 70	15 (3.9)	6 (3.1)	100 (62.5)	ANT-E *T* ≥ 78	7 (1.8)	3 (1.6)	78 (48.8)
ANT-S *T* ≥ 70	48 (12.5)	30 (15.6)	96 (60.0)	ANT-S *T* ≥ 81	19 (5.0)	8 (4.2)	52 (32.5)
AGG-A *T* ≥ 70	44 (11.5)	47 (24.5)	111 (69.4)	AGG-A *T* ≥ 81	8 (2.1)	16 (8.3)	49 (30.6)
AGG-V *T* ≥ 70	23 (6.0)	27 (14.1)	60 (37.5)	AGG-V *T* ≥ 74	17 (4.4)	18 (9.4)	51 (31.9)
AGG-P *T* ≥ 70	27 (7.0)	46 (24.0)	130 (81.3)	AGG-P *T* ≥ 93	4 (1.0)	9 (4.7)	72 (45.0)

Note. PAI: Personality Assessment Inventory; PTSD: Posttraumatic stress disorder; PTSD sim: PTSD simulators; See Supplemental Table 6 for subscale acronym definitions.

Table 4. Cumulative multivariate base rates of standard (left) and skyline (right) cutoff elevations on the PAI clinical scales.

Standard elevations				Skyline elevations			
	Group, *n* (%)				Group, *n* (%)		
Number of elevations	Mood dx *n* = 524	PTSD *n* = 252	PTSD sim *n* = 160	Number of elevations	Mood dx *n* = 524	PTSD *n* = 252	PTSD sim *n* = 160
0 (none)	198 (37.8)	37 (14.7)	12 (7.5)	0 (none)	473 (90.3)	170 (67.5)	36 (22.5)
1 or More	326 (62.2)	215 (85.3)	148 (92.5)	1 or More	51 (9.7)	82 (32.5)	124 (77.5)
2 or More	216 (41.2)	186 (73.8)	148 (92.5)	2 or More	14 (2.7)	40 (15.9)	104 (65.0)
3 or More	157 (30.0)	158 (62.7)	145 (90.6)	3 or More	3 (0.6)	23 (9.1)	93 (58.1)
4 or More	100 (19.1)	135 (53.6)	144 (90.0)	4 or More	1 (0.2)	15 (6.0)	84 (52.5)
5 or More	68 (13.0)	100 (39.7)	140 (87.5)	5 or More	1 (0.2)	8 (3.2)	69 (43.1)
6 or More	38 (7.3)	71 (28.2)	133 (83.1)	6 or More	1 (0.2)	3 (1.2)	57 (35.6)
7 or More	26 (5.0)	45 (17.9)	127 (79.4)	7 or More	1 (0.2)	2 (0.8)	46 (28.7)
8 or More	7 (1.3)	20 (7.9)	125 (78.1)	8 or More	0 (0.0)	2 (0.8)	37 (23.1)
9 or More	4 (0.8)	8 (3.2)	114 (71.3)	9 or More	0 (0.0)	0 (0.0)	27 (16.9)
10 or More	1 (0.2)	4 (1.6)	97 (60.6)	10 or More	0 (0.0)	0 (0.0)	14 (8.8)
11 (All)	1 (0.2)	1 (0.4)	31 (19.4)	11 (All)	0 (0.0)	0 (0.0)	4 (2.5)
j, Med (IQR)	1.0 (3.0)	4.0 (4.0)	10 (2.0)	*j, Med (IQR)*	0.0 (0.0)	0.0 (1.0)	4.0 (6.0)

Note. PAI: Personality Assessment Inventory; PTSD: posttraumatic stress disorder; PTSD sim: PTSD simulators; *j*: summed number of elevations; Med: median; IQR: interquartile range. Cells shaded gray indicate MBR of *j* elevations considered "unusual" (i.e., <10%).

ROC analyses differentiating PTSD and mood disorders from PTSD simulators

Supplemental Table 10 presents ROC curve summary statistics for MBR indicators, NIM, and MFI in differentiating (a) PTSD (referent) from PTSD simulators and (b) mood disorders (referent) from PTSD simulators (see Supplemental Figures 1 and 2 for respective ROC graphs). All AUC values yielded from MBR indicators and NIM were significant (all *p*-values < .001), and most fell in the "moderate" diagnostic efficacy interpretive range. Next, across models, standard-elevation MBR indicators were nominally superior to skyline-elevation ones, as were clinical scale relative to subscale. For the PTSD vs. PTSD simulator models, MBR AUC's ranged from .817 to .862 and were comparable to NIM and MFI (AUC = .864 and .818, respectively). These findings were mirrored in the mood disorders vs. PTSD simulator models, with MBR AUC's ranging from .866 to .915 and analogous to NIM and MFI (AUC = .913 and .912, respectively).

Table 5. Cumulative multivariate base rates of standard (left) and skyline (right) cutoff elevations on the PAI clinical subscales.

Standard elevations				Skyline elevations			
	Group, *n* (%)				Group, *n* (%)		
Number of elevations	Mood dx *n* = 383	PTSD *n* = 192	PTSD sim *n* = 160	Number of elevations	Mood dx *n* = 383	PTSD *n* = 192	PTSD sim *n* = 160
0 (none)	56 (14.6)	10 (5.2)	9 (5.6)	0 (none)	252 (65.8)	95 (49.5)	21 (13.1)
1 or More	327 (85.4)	182 (94.8)	151 (94.4)	1 or More	131 (34.2)	97 (50.5)	139 (86.9)
2 or More	278 (72.6)	169 (88.0)	149 (93.1)	2 or More	58 (15.1)	62 (32.3)	126 (78.8)
3 or More	239 (62.4)	158 (82.3)	148 (92.5)	3 or More	29 (7.6)	44 (22.9)	119 (74.4)
4 or More	196 (51.2)	148 (77.1)	147 (91.9)	4 or More	20 (5.2)	31 (16.1)	111 (69.4)
5 or More	169 (44.1)	138 (71.9)	145 (90.6)	5 or More	11 (2.9)	25 (13.0)	103 (64.4)
6 or More	140 (36.6)	136 (70.8)	145 (90.6)	6 or More	9 (2.3)	21 (10.9)	95 (59.4)
7 or More	119 (31.1)	128 (66.7)	144 (90.0)	7 or More	4 (1.0)	16 (8.3)	88 (55.0)
8 or More	98 (25.6)	113 (58.9)	143 (89.4)	8 or More	2 (0.5)	12 (6.3)	82 (51.2)
9 or More	83 (21.7)	100 (52.1)	141 (88.1)	9 or More	1 (0.3)	11 (5.7)	77 (48.1)
10 or More	67 (17.5)	92 (47.9)	137 (85.6)	10 or More	1 (0.3)	8 (4.2)	70 (43.8)
11 or More	51 (13.3)	85 (44.3)	137 (85.6)	11 or More	1 (0.3)	6 (3.1)	68 (42.5)
12 or More	34 (8.9)	76 (39.6)	134 (83.8)	12 or More	1 (0.3)	4 (2.1)	55 (34.4)
13 or More	25 (6.5)	60 (31.3)	132 (82.5)	13 or More	1 (0.3)	3 (1.6)	50 (31.3)
14 or More	17 (4.4)	48 (25.0)	129 (80.6)	14 or More	1 (0.3)	3 (1.6)	49 (30.6)
15 or More	15 (3.9)	45 (23.4)	126 (78.8)	15 or More	1 (0.3)	2 (1.0)	45 (28.1)
16 or More	10 (2.6)	38 (19.8)	122 (76.3)	16 or More	1 (0.3)	1 (0.5)	32 (20.0)
17 or More	6 (1.6)	28 (14.6)	119 (74.4)	17 or More	1 (0.3)	1 (0.5)	28 (17.5)
18 or More	5 (1.3)	20 (10.4)	115 (71.9)	18 or More	1 (0.3)	1 (0.5)	25 (15.6)
19 or More	4 (1.0)	15 (7.8)	108 (67.5)	19 or More	1 (0.3)	1 (0.5)	19 (11.9)
20 or More	2 (0.5)	11 (5.7)	98 (61.3)	20 or More	1 (0.3	1 (0.5)	14 (8.8)
21 or More	2 (0.5)	7 (3.6)	88 (55.0)	21 or More	0 (0.0)	0 (0.0)	10 (6.3)
22 or More	1 (0.3)	4 (2.1)	80 (50.0)	22 or More	0 (0.0)	0 (0.0)	6 (3.8)
23 or More	1 (0.3)	3 (1.6)	72 (45.0)	23 or More	0 (0.0)	0 (0.0)	3 (1.9)
24 or More	0 (0.0)	1 (0.5)	67 (41.9)	24 or More	0 (0.0)	0 (0.0)	1 (0.6)
25 or More	0 (0.0)	1 (0.5)	49 (30.6)	25 or More	0 (0.0)	0 (0.0)	0 (0.0)
26 or More	0 (0.0)	1 (0.5)	29 (18.1)	26 or More	0 (0.0)	0 (0.0)	0 (0.0)
27 or More	0 (0.0)	0 (0.0)	12 (7.5)	27 or More	0 (0.0)	0 (0.0)	0 (0.0)
28 (All)	0 (0.0)	0 (0.0)	1 (0.6)	28 (All)	0 (0.0)	0 (0.0)	0 (0.0)
j, Med (IQR)	4.0 (7.0)	9 (9.8)	21.5 (9.0)	*j, Med (IQR)*	0.0 (1.0)	1.0 (2.0)	8.0 (13.0)

Note. PAI: Personality Assessment Inventory; PTSD: Posttraumatic stress disorder; PTSD sim: PTSD simulators; *j*: summed number of elevations. Cells shaded gray indicate MBR of *j* elevations considered "unusual" (i.e., <10%).

Optimal cutoffs (≥) of *j* identified by the Youden's index for each MBR marker approximated *j* thresholds identified using Brooks et al. (2011) definition of *unusual* MBR (i.e., < 10%). For example, the *unusual* MBR thresholds on PAI clinical scales for the PTSD group were $j \geq 8$ for standard elevations and $j \geq 3$ for skyline elevations. These were near identical to the Youden's index cutoffs in the PTSD vs. PTSD simulators models: $j \geq 7.5$ for standard elevations and $j \geq 1.5$ for skyline elevations on PAI clinical scales. Importantly, MBR of standard elevations on clinical scales yielded the highest Youden's statistic (*J*) compared to all other variables in both models, indicating strongest diagnostic performance.

Irrespective of referent (i.e., PTSD or mood disorders), AUC analysis indicated the optimal MBR cutoffs yielded acceptable sensitivity (.650 to .806) and robust specificity (.833 to .984) classification statistics. The NIM and MFI recommended cutoffs produced similar sensitivity (.650 to .763) and specificity (.771 to .990) statistics. Next, for MBR-based cutoffs, overall classification accuracy ranged from 76.7 to 86.7% for discerning PTSD from PTSD simulators, and 87.1 to 93.9% for discerning mood disorders from PTSD simulators. These markers performed similarly to the recommended NIM and MFI cutoffs, which correctly classified 83.0% and 77.9% of cases (respectively) for the PTSD vs. PTSD simulators model, and 90.8% and 92.0% of cases (respectively) for the mood disorders vs. PTSD simulators model. Lastly, substantial agreement was observed among all indicators (κ'*s* = .652 to .914), irrespective of model, and there was a trend of stronger agreement between MBR and MFI cutoffs (see Supplemental Table 11). We conducted sensitivity analyses including the PTSD simulators ($n = 55$) who were originally excluded for ICN ≥ 73 and/or INF ≥ 75. All ROC models and classification rates remained consistent with the initial analyses with AUC value Δ'*s* ≤ .010 (see Supplemental Table 12). Thus, MBR showed utility in detecting simulators with systematic and/or unsystematic distortion.

As a final exploratory analysis, we tested the incremental validity of MBR versus NIM and MFI in predicting simulated PTSD. We entered each MBR variable (standard scale, skyline subscale, standard subscale, skyline subscale) in separate multinomial logistic regression models alongside NIM and MFI to predict group (genuine PTSD, mood disorders, PTSD simulators) with PTSD simulators set as the referent. For ease of interpretability, we reran models setting comparison groups, e.g., genuine PTSD, as the referent so that positive odds ratios indicate increased odds of PTSD simulator group status for a given variable. MBRs were incrementally predictive of group status across virtually all models when accounting for NIM and MFI. Standard elevations on PAI clinical scales demonstrated the strongest incremental effect for predicting PTSD simulator group status compared to genuine PTSD (OR = 1.90; 95% CI = 1.54, 2.33; $p < .001$) and mood disorders (OR = 1.89; 95% CI = 1.52, 2.34; $p < .001$). That is, profiles with more elevations were associated with increased odds of simulated PTSD relative to examinees in our genuine PTSD and mood disorders groups.

NIM was significant across models but with smaller incremental effects (ORs ranging from 1.06 to 1.09). Higher MFI was generally not associated with increased odds of simulated PTSD group membership after adjusting for other indicators (see Supplemental Table 13).

Discussion

Review of key findings

This study presents comprehensive *univariate* base rates (UBR) and *multivariate* base rates (MBR) of standard and skyline-elevations across PAI indicators (clinical scales and subscales), and for multiple groups (mood disorders, genuine PTSD, and PTSD simulators). We also evaluated the diagnostic efficacy of MBR outcomes, which yielded broadly acceptable sensitivity and specificity values for clinical use (Larrabee et al., 2019), which were comparable to those from NIM. UBR on most PAI indicators were highest in PTSD simulators followed by the PTSD and then mood disorders group. This same probability gradient (PTSD simulators > PTSD > mood disorders) was observed across MBR outcomes. Our data show compelling evidence that it is not uncommon for examinees with PTSD and mood disorder diagnoses to obtain *some* standard and skyline elevations on the PAI. Critically, marked differences in MBR were observed in the mood and PTSD groups vs. the PTSD simulators. Exploratory analyses also revealed superior incremental validity of MBRs (especially number of standard elevations across clinical scales) in predicting simulated PTSD (from genuine PTSD or mood disorders) when adjusting for NIM and MFI. However, we advise cautious interpretation of those findings given the inter-correlations and overlap of items among these embedded PAI indicators.

Synthesis of findings with prior literature

Keeping with past research on negative distortion PAI scales, MBR indicators significantly differed between examinees with genuine vs. feigned PTSD presentations. While prior literature found Cohen's *d* values ranging between 0.60 and 1.60 for the existing scales (Calhoun et al., 2010; Liljequist et al., 1998; Rogers et al., 1993; Russell & Morey, 2019; Scragg et al., 2000; Thomas et al., 2012), MBR indicator effects consistently exceeded 1.00 (*d* ranging from −1.33 to −1.65) and were comparable to NIM ($d = -1.78$) and MFI ($d = 1.30$) in our data (see Supplemental Table 14). The MBR indicators also demonstrated comparable diagnostic accuracy to the most robust findings in past literature (which is with variable findings for extant scales). For example, the strongest reported findings for NDS captured 97% of genuine PTSD and 64% of feigned PTSD cases (Thomas et al., 2012), which is comparable to our classification rates for the identified MBR cutoff for clinical scale standard elevations (92.1% of genuine PTSD and 78.1% of simulated PTSD cases correctly identified). The clinical scale standard elevation MBR sensitivity and specificity statistics (.781 and .917, respectively) are also consistent with the most robust MFI findings reported (i.e., .591 and .920, respectively; Russell & Morey, 2019). The comparable, if not stronger, findings in our study relative to past research should be contextualized in the setting of methodological differences across studies, namely that we relied on a simulator sample whereas other studies featured examinees feigning under natural circumstances.

Over-reporting detection on broadband personality measures involves a variety of theorized methods as a basis for over-reporting scale development (e.g., rare, quasi-rare; Burchett & Bagby, 2022; Rogers & Bender, 2020) and individual scales (e.g., NIM) generally produce effects within a stable and standard, general range across each of the classification methods used, regardless of instrument. On the PAI, for instance, meta-analytic findings for over-reporting scales have consistently evidenced (1) a general effect (*d*) range of 0.75 to 1.50 for standardized mean differences that vary systematically according to the evaluation context, (2) higher standard deviations in identified feigning groups (both criterion and simulation design), up to approximately twice the size of normative data, and (3) sensitivities between .10 and .50 when specificity is set to .90 or greater (Hawes & Boccaccini, 2009; Herring et al., 2025). Indeed, various supplemental personality assessment scales, including those on the PAI (e.g., MFI; Gaines et al., 2013), rely on this broad response tendency to detect feigned presentations, as does research demonstrating regular co-elevation even amongst distinct domain focused scales (e.g., Ingram et al., 2020). Even when scales are refined and narrowed in their construction to have only "face valid" content of a specific domain (e.g., cognitive), those revised SVTs tend to produce statistically equivalent effects to scales prior to revision (see Ingram et al., 2024; Ratcliffe et al., 2024).

This tendency of over-reporting to occur as a general "saturated" response comes in contrast with practice-based recommendations to differentiate SVT presentations according to cognitive, somatic, and psychological symptoms (e.g., Rogers & Bender, 2020; Sherman et al., 2020). This generalized response could be conceptualized as a "Symptom Severity" approach, given that elevations on such substantive scales, ipso facto, indicate a higher degree of symptoms. However, symptom severity approaches, while largely under-developed within the SVT literature, rely on a diversity of symptom domains (Rogers & Bender, 2020). In contrast, at both univariate and multivariate levels, the diversity of symptom domains appears to matter minimally. In this vein, Gaines et al. (2013) noted that "Individuals tend to report numerous symptoms across a wide variety of diagnostic categories rather than focusing on a particular diagnostic category" and that those individuals "tend to report that most, if not all, of their symptoms are quite severe rather than mild or moderate" (p. 439). Likewise, Ingram et al. (2024) conclude their approach works because of the underlying "infrequency-based" methods wherein "summed or otherwise parceled information (e.g., substantive scale scores)", and reference MBR as an example of this detection pattern (p. 13). Thus, generalized responding is not dependent on symptom diversity and MBR's effectiveness and strength comes from the tendency for those distinctions not to matter, as they do not show up in a high symptom endorsement profile.

Accordingly, the MBR approach provides another method to make determinations of over-reporting using elevation ceilings and fits within the recent proliferation of symptom validity-focused research and methodologic advancements (Suchy, 2019; Whiteside & Basso, 2024). Given the low frequency of a ceiling elevation (i.e., skyline profile on the PAI), as well as the even lower cumulative frequency of multiple such elevations, the MBR approach offers an additional indicator of over-reported, invalid responding. MBR also offer a warning sign of subsequent poor assessment data quality, as would any elevated validity scale. Such a finding, however, does not negate the usefulness of other such scales. NIM, for instance, has also produced strong, standardized effect sizes (Hawes & Boccaccini, 2009; Herring et al., 2025). Further, NIM and MFI performed well (albeit slightly less so than MBR) in the current study, and agreement among respective cutoffs was substantial. These trends likely reflect the high portion of shared variance across SVTs (Shura et al., 2025). Notably, however, the MBR approach tends to produce higher classification estimates, particularly concerning sensitivity (Larrabee, 2012).

At present, detection of SVT-based over-reporting should bear in mind that the conclusions for when a scale is elevated rests on specificity (e.g., low rate of false positives; Sherman et al., 2020) and that high specificity does not indicate a true positive, per se (e.g., we are not concluding that an individual over-reports necessarily). If replicated across a diversity of criterion-identified studies, these findings may lead to a preference for MBR-based measures given this strength in scale sensitivity. It is also worth noting that over-reporting classification methods and statistics have also faced cogent criticisms (e.g., Leonhard, 2023) though not all agree with those criticisms (e.g., Bush, 2023; Young & Erdodi, 2024). This study's support for MBR principles suggests a need to empirically test domain-specific SVT practice guidelines (Sherman et al., 2020; Sweet et al., 2021).

Limitations

While this study possessed several strengths such as our robust well-defined clinical samples, comprehensive approach to appraising likelihoods of clinical scale and subscale elevations on the PAI, and considering skyline elevation cutoffs for each protocol analyzed, several limitations warrant discussion. First, sex and age differences were observed among groups, particularly the simulator group (which was largely composed of females in their 20s). It is possible that simulator groups with a different demographic composition may *simulate* PTSD differently. In a somewhat similar vein, the mood disorders and PTSD groups were largely comprised of older males. More broadly, most of the study population identified as non-Hispanic White and this may limit generalizability of findings. Encouragingly, follow-up within-group analyses revealed that demographic factors were not meaningfully related to MBR outcomes. This is in-line with prior MBR literature on self-report instruments (Aita et al., 2022; 2023; Ingram & Karr, 2024). Nevertheless, future lines of research should consider replicating our findings in demographically matched, and racially/ethnically diverse, groups.

Another important limitation is our lack of detailed information with respect to PTSD diagnoses such as data on inter-rater reliability of diagnoses, specific diagnostic methods employed (e.g., structured and/or unstructured interviews), and how final determinations of whether participants met diagnostic criteria were adjudicated. This limitation is partly attributable to the complex series of diverse subsamples we drew from applied settings (some of which included retrospective chart reviews and reliance on historical diagnoses on problem lists) to address our research aims. Nonetheless, this limitation potentially challenges the accuracy of psychiatric diagnoses for our clinical groups. Relatedly, we only featured one general PTSD group and do not account for distinct PTSD presentations given the disorder's symptom heterogeneity (Borgogna et al., 2024) although research with the PAI may not be aligned to do so given the lack of profile discrimination (Ingram et al., 2021). While we argue that MBR approach would likely demonstrate comparable diagnostic utility across clinical presentations as it makes use of all PAI information, future studies are necessary to clarify the generalizability of our findings to PTSD subtypes and other psychiatric conditions. In a different vein, we were only able to examine NIM and MFI alongside MBR due to differences in available SVTs in the distinct datasets we combined for this study. Future research examining an expanded selection of PAI SVTs in relation to MBR is warranted.

Next, our PTSD group was exclusively composed of respondents from military contexts (i.e., Veterans and SMs) who largely had combat-related trauma experiences. As such, it is unclear whether MBR from our military-based PTSD group generalizes to civilians with PTSD. Indeed, literature indicates trauma type and other sociodemographic variables are associated with PTSD presentation (Hetzel-Riggin & Roby, 2013; Xue et al., 2015). Future studies featuring civilian respondents with PTSD and other various psychiatric disorders will be crucial to clarify the generalizability of the MBR data we present here. Finally, we relied on PTSD simulators participating for class credit to serve as our target invalidity group (criterion). Moreover, we did not provide clear definitions of "trauma" and additional context regarding the nature of the trauma to be feigned (e.g., combat-related). As such, our simulator group may not generalize to other contexts such as VA compensation claimants. It is unclear whether our simulator group is a valid proxy for actual respondents with financial incentives who are feigning PTSD. Importantly, simulator groups are often used in PVT and SVT research (Aita et al., 2018; Brennan et al., 2009; Lange et al., 2013; Lau et al., 2017), though that fact alone does not directly address the issue of equivalency between simulators and actual feigners. Prior literature indicates simulation-based validity studies overestimate sensitivity while underestimating specificity rates in the criterion group (Sherman et al., 2020). Further, simulator coaching that incorporates external incentives (in-line with our procedure) yields more exaggerated levels of symptom endorsement (Rogers & Cruise, 1998).

Finally, base rates of malingering vary widely depending on clinical setting and population (Aita et al., 2020; Mittenberg et al., 2002). Our base rate of simulators within the study population ranged from 23% to 39%, across analyses, which in turn, has implications for future efforts to replicate our calculated diagnostic accuracy estimates. Moreover, a test's sensitivity and specificity are differentially susceptible to error posed by low and high base rate conditions, respectively (Robinson et al., 2016). However, it should be noted that our base rate simulator numbers are lower than the base rate of failed validity indicators generally observed in medicolegal contexts (Larrabee et al., 2009).

Given these possible limitations, additional studies are needed to replicate our findings, but the current results demonstrate the utility of using an MBR methodology to detect symptom exaggeration on broad measures of emotional functioning such as the PAI. If replicated, the MBR approach may serve as an additional supplemental indicator of protocol validity for the PAI (e.g., McCredie & Morey, 2018). Applying MBR probabilities enhances interpretive accuracy when considering an entire PAI response profile. Tallying the number of elevated scales or subscales a respondent produces helps differentiate genuine PTSD as well as mood disorders from feigned PTSD. We advise clinicians and researchers use multiple validated approaches when appraising credibility of examinee presentations (Sherman et al., 2020; Wygant et al., 2007).

Acknowledgments

This research reported/outlined here was supported by the Department of Veterans Affairs, Veterans Health Administration, VISN 1 Career Development Award to Stephen L. Aita (V1CDA2024-74). The other authors do not have any funding sources to report. The authors would like to acknowledge Jennifer Greene and Sierra Iwanicki for their assistance with data acquisition. The views, opinions, and/or findings contained in this article are those of the authors and should not be construed as an official Defense Health Agency, Department of Veterans Affairs, or U.S. Government position, policy, or decision unless so designated by other official documentation.

Declaration of interest

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Robert M. Roth is an author of the BRIEF-A and receives royalties from the publisher of the PAI (Psychological Assessment Resources, Inc. [PAR]). Paul B. Ingram is paid by PAR Inc. as an advisory board member for the PAI. He also receives research support from the University of Minnesota Press in the form of Minnesota Multiphasic Personality Inventory test administrations. The other authors declare no conflicts of interest.

ORCID

Stephen L. Aita http://orcid.org/0000-0001-7201-3247
Louis A. Pagano Jr. http://orcid.org/0009-0007-9314-2356
Michael J. Broggi http://orcid.org/0000-0002-8402-0329
Paul B. Ingram http://orcid.org/0000-0002-5409-4896
Nicholas C. Borgogna http://orcid.org/0009-0007-4436-8500
Grant G. Moncrief http://orcid.org/0000-0001-5811-4308
Robert M. Roth http://orcid.org/0000-0003-4374-6569
Matthew R. Calamia http://orcid.org/0000-0002-7252-7181
Patrick Armistead-Jehle http://orcid.org/0000-0002-6784-1432
Benjamin D. Hill http://orcid.org/0000-0002-5797-5082

Data availability statement

The data that support the findings of this study are property of the U.S. Government and are not available upon request.

References

Aita, S. L., Borgogna, N. C., Aita, L. J., Ogden, M. L., & Hill, B. D. (2020). Comparison of clinical psychologist and physician beliefs and practices concerning malingering: Results from a mixed methods study. *Psychological Injury and Law*, *13*(3), 246–260. https://doi.org/10.1007/s12207-020-09374-x

Aita, S. L., Hill, B. D., Musso, M. W., & Gouvier, W. D. (2018). Can we identify bad cops based on history? Base rates of historical markers in law enforcement pre-employment evaluations. *Journal of Police and Criminal Psychology*, *33*(3), 201–208. https://doi.org/10.1007/s11896-018-9258-0

Aita, S. L., Holding, E. Z., Greene, J., Carrillo, A., Moncrief, G. G., Isquith, P. K., Gioia, G. A., & Roth, R. M. (2022). Multivariate base rates of score elevations on the BRIEF2 in children with ADHD, autism spectrum disorder, or specific learning disorder with impairment in reading. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*, *28*(7), 979–996. https://doi.org/10.1080/09297049.2022.2060201

Aita, S. L., Moncrief, G. G., Carrillo, A., Greene, J., Trujillo, S., Gioia, G. A., Isquith, P. K., & Roth, R. M. (2022). Enhanced interpretation of the BRIEF2: Multivariate base rates of elevated scores in the standardization samples. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*, *28*(4), 535–553. https://doi.org/10.1080/09297049.2021.1998408

Aita, S. L., Moncrief, G. G., Greene, J., Trujillo, S., Carrillo, A., Iwanicki, S., Morera, C. C., Gioia, G. A., Isquith, P. K., & Roth, R. M. (2023). Univariate and multivariate base rates of score elevations, reliable change, and inter-rater discrepancies in the BRIEF-A standardization samples. *Assessment*, *30*(2), 390–401. https://doi.org/10.1177/10731911211055673

Aita, S. L., Sofko, C. A., Hill, B. D., Musso, M. W., & Boettcher, A. C. (2018). Utility of the personality assessment inventory in detecting feigned attention-deficit/hyperactivity disorder (ADHD): The feigned adult ADHD index. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *33*(7), 832–844. https://doi.org/10.1093/arclin/acx113

American Psychiatric Association. (2000). Diagnostic and statistical manual of mental disorders (4th ed.), *Text Revision (DSM-IV-TR). Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR)*. https://doi.org/10.1176/APPI.BOOKS.9780890423349

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.).

Baratloo, A., Hosseini, M., Negida, A., & Ashal, G. E. (2015). Part 1: Simple definition and calculation of accuracy, sensitivity and specificity. *Emergency (Tehran, Iran)*, *3*(2), 48–49.

Borgogna, N. C., Lathan, E. C., & Aita, S. L. (2023). Sexual and gender minority victimization: Base rates of assault in college students across sexual and gender identities. *Journal of Interpersonal Violence*, *38*(7-8), 5613–5637. https://doi.org/10.1177/08862605221124253

Borgogna, N. C., Owen, T., & Aita, S. L. (2024). The absurdity of the latent disease model in mental health: 10,130,814 ways to have a DSM-5-TR psychological disorder. *Journal of Mental Health (Abingdon, England)*, *33*(4), 451–459. https://doi.org/10.1080/09638237.2023.2278107

Brennan, A. M., Meyer, S., David, E., Pella, R., Hill, B. D., & Gouvier, W. D. (2009). The vulnerability to coaching across measures of ef-

fort. *The Clinical Neuropsychologist*, *23*(2), 314–328. https://doi.org/10.1080/13854040802054151

Brooks, B. L., Holdnack, J. A., & Iverson, G. L. (2011). Advanced clinical interpretation of the WAIS-IV and WMS-IV: Prevalence of low scores varies by level of intelligence and years of education. *Assessment*, *18*(2), 156–167. https://doi.org/10.1177/1073191110385316

Brooks, B. L., Iverson, G. L., & Holdnack, J. A. (2013). Understanding and using multivariate base rates with the WAIS–IV/WMS–IV. In J. A. Holdnack, L. W. Drozdick, L. G. Weiss, & G. L. Iverson (Eds.), *WAIS-IV, WMS-IV, and ACS: Advanced clinical interpretation* (pp. 75–102). Elsevier Science. https://doi.org/10.1016/B978-0-12-386934-0.00002-X

Burchett, D., & Bagby, R. M. (2022). Assessing negative response bias: A review of the noncredible overreporting scales of the MMPI-2-RF and MMPI-3. *Psychological Injury and Law*, *15*(1), 22–36. https://doi.org/10.1007/s12207-021-09435-9

Bush, S. S. (2023). Questioning what we thought we knew: Commentary on Leonhard's performance validity assessment articles. *Neuropsychology Review*, *33*(3), 624–627. https://doi.org/10.1007/S11065-023-09603-5

Calhoun, P. S., Collie, C. F., Clancy, C. P., Braxton, L. E., & Beckham, J. C. (2010). Use of the PAI in assessment of posttraumatic stress disorder among help-seeking veterans. In M. A. Blais, M. R. Baity, & C. J. Hopwood (Eds.), *Clinical applications of the personality assessment inventory* (1st ed., pp. 93–111). Taylor & Francis. https://doi.org/10.4324/9780203856604-11

Calhoun, P. S., Earnst, K. S., Tucker, D. D., Kirby, A. C., & Beckham, J. C. (2000). Feigning combat-related posttraumatic stress disorder on the personality assessment inventory. *Journal of Personality Assessment*, *75*(2), 338–350. https://doi.org/10.1207/S15327752JPA7502_11

Crawford, J. R., Garthwaite, P. H., Sutherland, D., & Borland, N. (2011). Some supplementary methods for the analysis of the Delis-Kaplan Executive Function System. *Psychological Assessment*, *23*(4), 888–898. https://doi.org/10.1037/a0023712

Department of Veterans Affairs (2015). *Analysis of VA health care utilization among Operation Enduring Freedom (OEF), Operation Iraqi Freedom (OIF), and Operation New Dawn (OND) veterans*. https://www.publichealth.va.gov/docs/epidemiology/healthcare-utilization-report-fy2015-qtr2.pdf

Eisenberg, M. J. (1995). Accuracy and predictive values in clinical decision-making. *Cleveland Clinic Journal of Medicine*, *62*(5), 311–316. https://doi.org/10.3949/ccjm.62.5.311

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, *40*(5), 532–538. https://doi.org/10.1037/a0015808

Freeman, T., Powell, M., & Kimbrell, T. (2008). Measuring symptom exaggeration in veterans with chronic posttraumatic stress disorder. *Psychiatry Research*, *158*(3), 374–380. https://doi.org/10.1016/j.psychres.2007.04.002

Gaines, M. V., Giles, C. L., & Morgan, R. D. (2013). The detection of feigning using multiple PAI scale elevations. *Assessment*, *20*(4), 437–447. https://doi.org/10.1177/1073191112458146

Gervais, R. O., Russell, A. S., Green, P., Allen, L. M., Ferrari, R., & Pieschl, S. D. (2001). Effort testing in patients with fibromyalgia and disability incentives. *The Journal of Rheumatology*, *28*(8), 1892–1899.

Gioia, G. A., Isquith, P. K., Guy, S. C., Kenworthy, L. (2015). *BRIEF2: Behavior Rating Inventory of Executive Function, Second Edition professional manual*. Psychological Assessment Resources. https://www.parinc.com/products/pkey/23

Green, P., Rohling, M. L., Lees-Haley, P. R., & Allen, L. M. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury*, *15*(12), 1045–1060. https://doi.org/10.1080/02699050110088254

Hawes, S. W., & Boccaccini, M. T. (2009). Detection of overreporting of psychopathology on the Personality Assessment Inventory: A meta-analytic review. *Psychological Assessment*, *21*(1), 112–124. https://doi.org/10.1037/A0015036

Herring, T. T., Albertorio, A. D., Diehl, K. J., & Ingram, P. B. (2025). A meta-analysis of the Personality Assessment Inventory (PAI) over-reporting scales and supplemental indicators. *Journal of Psychopathology and Behavioral Assessment*. https://doi.org/10.1007/s10862-025-10233-9

Hetzel-Riggin, M. D., & Roby, R. P. (2013). Trauma type and gender effects on PTSD, general distress, and peritraumatic dissociation. *Journal of Loss and Trauma*, *18*(1), 41–53. https://doi.org/10.1080/15325024.2012.679119;SUBPAGE:STRING:FULL

Hill, B. D., Womble, M. N., & Rohling, M. L. (2015). Logistic regression function for detection of suspicious performance during baseline evaluations using Concussion Vital Signs. *Applied Neuropsychology. Adult*, *22*(3), 233–240. https://doi.org/10.1080/23279095.2014.910215

Hong, S. H., & Kim, Y. H. (2001). Detection of random response and impression management in the PAI: II. Detection indices. *Korean Journal of Clinical Psychology*, *20*(4), 751–761. https://accesson.kr/kjcp/v.20/4/751/26761

Ingram, E. O., & Karr, J. E. (2024). Multivariate base rates of potentially problematic scores on the NIH Toolbox Emotion Battery. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *39*(4), 454–463. https://doi.org/10.1093/ARCLIN/ACAD094

Ingram, P. B., Armistead-Jehle, P., Childers, L. G., & Herring, T. T. (2024). Cross validation of the response bias scale and the response bias scale-19 in active-duty personnel: Use on the MMPI-2-RF and MMPI-3. *Journal of Clinical and Experimental Neuropsychology*, *46*(2), 141–151. https://doi.org/10.1080/13803395.2024.2330727

Ingram, P. B., Sharpnack, J. D., Mosier, N. J., & Golden, B. L. (2021). Evaluating symptom endorsement typographies of trauma-exposed veterans on the Personality Assessment Inventory (PAI): A latent profile analysis. *Current Psychology*, *40*(11), 5267–5277. https://doi.org/10.1007/S12144-019-00486-5/METRICS

Ingram, P. B., Tarescavage, A. M., Ben-Porath, Y. S., & Oehlert, M. E. (2020). Patterns of MMPI-2-Restructured Form (MMPI-2-RF) validity scale scores observed across Veteran Affairs settings. *Psychological Services*, *17*(3), 355–362. https://doi.org/10.1037/SER0000339

Karr, J. E., Garcia-Barrera, M. A., Holdnack, J. A., & Iverson, G. L. (2017). Using multivariate base rates to interpret low scores on an abbreviated battery of the Delis-Kaplan Executive Function System. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *32*(3), 297–305. https://doi.org/10.1093/arclin/acw105

Karr, J. E., Garcia-Barrera, M. A., Holdnack, J. A., & Iverson, G. L. (2018). Advanced clinical interpretation of the Delis-Kaplan Executive Function System: Multivariate base rates of low scores. *The Clinical Neuropsychologist*, *32*(1), 42–53. https://doi.org/10.1080/13854046.2017.1334828

Labarge, A. S., McCaffrey, R. J., & Brown, T. A. (2003). Neuropsychologists' abilities to determine the predictive value of diagnostic tests. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *18*(2), 165–175. https://doi.org/10.1016/S0887-6177(01)00193-7

Landis, J. R., & Koch, G. G. (1977). The Measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Lange, R. T., Edmed, S. L., Sullivan, K. A., French, L. M., & Cooper, D. B. (2013). Utility of the mild brain injury atypical symptoms scale to detect symptom exaggeration: An analogue simulation study. *Journal of Clinical and Experimental Neuropsychology*, *35*(2), 192–209. https://doi.org/10.1080/13803395.2012.761677

Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society: JINS*, *18*(4), 625–630. https://doi.org/10.1017/S1355617712000240

Larrabee, G. J., Millis, S. R., & Meyers, J. E. (2009). 40 plus or minus 10, a new magical number: Reply to Russell. *The Clinical Neuropsychologist*, *23*(5), 841–849. https://doi.org/10.1080/13854040902796735

Larrabee, G. J., Rohling, M. L., & Meyers, J. E. (2019). Use of multiple performance and symptom validity measures: Determining the optimal per test cutoff for determination of invalidity, analysis of skew, and inter-test correlations in valid and invalid performance groups. *The Clinical Neuropsychologist*, *33*(8), 1354–1372. https://doi.org/10.1080/13854046.2019.1614227

Lau, L., Basso, M. R., Estevis, E., Miller, A., Whiteside, D. M., Combs, D., & Arentsen, T. J. (2017). Detecting coached neuropsychological

dysfunction: A simulation experiment regarding mild traumatic brain injury. *The Clinical Neuropsychologist*, *31*(8), 1412–1431. https://doi.org/10.1080/13854046.2017.1318954

Lees-Haley, P. R. (1997). MMPI-2 base rates for 492 personal injury plaintiffs: Implications and challenges for forensic assessment. *Journal of Clinical Psychology*, *53*(7), 745–755. https://doi.org/10.1002/(SICI)1097-4679(199711)53:7

Leonhard, C. (2023). *Quo Vadis* forensic neuropsychological malingering determinations? Reply to Drs. Bush, Faust, and Jewsbury. *Neuropsychology Review*, *33*(3), 653–657. https://doi.org/10.1007/S11065-023-09606-2

Liljequist, L., Kinder, B. N., & Schinka, J. A. (1998). An investigation of malingering posttraumatic stress disorder on the Personality Assessment Inventory. *Journal of Personality Assessment*, *71*(3), 322–336. https://doi.org/10.1207/S15327752JPA7103_3

McCredie, M. N., & Morey, L. C. (2018). Evaluating new supplemental indicators for the Personality Assessment Inventory: Standardization and cross-validation. *Psychological Assessment*, *30*(10), 1292–1299. https://doi.org/10.1037/PAS0000574

Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin*, *52*(3), 194–216. https://doi.org/10.1037/H0048070

Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, *24*(8), 1094–1102. https://doi.org/10.1076/JCEN.24.8.1094.8379

Mogge, N. L., Lepage, J. S., Bell, T., & Ragatz, L. (2010). The negative distortion scale: A new PAI validity scale. *Journal of Forensic Psychiatry & Psychology*, *21*(1), 77–90. https://doi.org/10.1080/14789940903174253

Morey, L. C. (2007). *Personality Assessment Inventory (PAI): Professional manual*. Psychological Assessment Resources.

Mozley, S. L., Miller, M. W., Weathers, F. W., Beckham, J. C., & Feldman, M. E. (2005). Personality Assessment Inventory (PAI) profiles of male veterans with combat-related posttraumatic stress disorder. *Journal of Psychopathology and Behavioral Assessment*, *27*(3), 179–189. https://doi.org/10.1007/s10862-005-0634-6

Musso, M. W., Hill, B. D., Barker, A. A., Pella, R. D., & Gouvier, W. D. (2016). Utility of the personality assessment inventory for detecting malingered ADHD in college students. *Journal of Attention Disorders*, *20*(9), 763–774. https://doi.org/10.1177/1087054714548031

Oltra-Cucarella, J., Sánchez-SanSegundo, M., Rubio-Aparicio, M., Arango-Lasprilla, J. C., & Ferrer-Cascales, R. (2021). The association between the number of neuropsychological measures and the base rate of low scores. *Assessment*, *28*(3), 955–963. https://doi.org/10.1177/1073191119864646

Palmer, B. W., Boone, K. B., Lesser, I. M., & Wohl, M. A. (1998). Base rates of "impaired" neuropsychological test performance among healthy older adults. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *13*(6), 503–511. https://doi.org/10.1093/arclin/13.6.503

Proto, D. A., Pastorek, N. J., Miller, B. I., Romesser, J. M., Sim, A. H., & Linck, J. F. (2014). The dangers of failing one or more performance validity tests in individuals claiming mild traumatic brain injury-related postconcussive symptoms. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *29*(7), 614–624. https://doi.org/10.1093/ARCLIN/ACU044

Ratcliffe, L. N., Hale, A. C., Gradwohl, B. D., & Spencer, R. J. (2024). Preliminary findings from reevaluating the MMPI Response Bias Scale items in veterans undergoing neuropsychological evaluation. *Applied Neuropsychology. Adult*, *31*(5), 1016–1023. https://doi.org/10.1080/23279095.2022.2106571

Robinson, M., Boissoneault, J., Sevel, L., Letzen, J., & Staud, R. (2016). The effect of base rate on the predictive value of brain biomarkers. *The Journal of Pain*, *17*(6), 637–641. https://doi.org/10.1016/j.jpain.2016.01.476

Rogers, R., & Bender, S. D. (2020). *Clinical assessment of malingering and deception* (4th ed.) The Guilford Press.

Rogers, R., & Cruise, K. R. (1998). Assessment of malingering with simulation designs: Threats to external validity. *Law and Human Behavior*, *22*(3), 273–285. https://doi.org/10.1023/A:1025702405865

Rogers, R., Ornduff, S. R., & Sewell, K. W. (1993). Feigning specific disorders: A study of the Personality Assessment Inventory (PAI). *Journal of Personality Assessment*, *60*(3), 554–560. https://doi.org/10.1207/S15327752JPA6003_12

Rogers, R., Sewell, K. W., Morey, L. C., & Ustad, K. L. (1996). Detection of feigned mental disorders on the Personality Assessment Inventory: A discriminant analysis. *Journal of Personality Assessment*, *67*(3), 629–640. https://doi.org/10.1207/s15327752jpa6703_15

Roth, R. M., Isquith, P. K., & Gioia, G. A. (2005). *Behavior rating inventory of executive function-adult version: Professional manual*. Psychological Assessment Resources.

Russell, D. N., & Morey, L. C. (2019). Use of validity indicators on the personality assessment inventory to detect feigning of post-traumatic stress disorder. *Psychological Injury and Law*, *12*(3-4), 204–211. https://doi.org/10.1007/S12207-019-09349-7/METRICS

Scragg, P., Bor, R., & Mendham, M.-C. (2000). Feigning post-traumatic stress disorder on the PAI. *Clinical Psychology & Psychotherapy*, *7*(2), 155–160. https://doi.org/10.1002/(SICI)1099-0879(200005)7:2<155::AID-CPP237>3.0.CO;2-Z

Seal, K. H., Metzler, T. J., Gima, K. S., Bertenthal, D., Maguen, S., & Marmar, C. R. (2009). Trends and risk factors for mental health diagnoses among Iraq and Afghanistan veterans using department of Veterans Affairs Health Care, 2002-2008. *American Journal of Public Health*, *99*(9), 1651–1658. https://doi.org/10.2105/AJPH.2008.150284

Sherman, E. M. S., Slick, D. J., & Iverson, G. L. (2020). Multidimensional malingering criteria for neuropsychological assessment: A 20-year update of the malingered neuropsychological dysfunction criteria. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *35*(6), 735–764. https://doi.org/10.1093/arclin/acaa019

Shura, R. D., Richard, K. W., Martindale, S. L., Brearly, T. W., Taber, K. H., & Canu, W. H. (2022). Internalizing and externalizing comorbidity and symptom burden in a VA ADHD specialty evaluation clinic. *Psychiatry Research*, *309*, 114395. https://doi.org/10.1016/J.PSYCHRES.2022.114395

Shura, R., Ingram, P. B., Schroder, R., & Armistead-Jehle, P. (2025). Interpreting the Personality Assessment Inventory (PAI) validity scales in Veteran Affairs (VA): Leveraging a longitudinal national sample for large data analytics. *Psychological Assessment*.

Suchy, Y. (2019). Introduction to special issue: Current trends in empirical examinations of performance and symptom validity. *The Clinical Neuropsychologist*, *33*(8), 1349–1353. https://doi.org/10.1080/13854046.2019.1672334

Sweet, J. J., Heilbronner, R. L., Morgan, J. E., Larrabee, G. J., Rohling, M. L., Boone, K. B., Kirkwood, M. W., Schroeder, R. W., Suhr, J. A. (2021). American Academy of Clinical Neuropsychology (AACN) 2021 consensus statement on validity assessment: Update of the 2009 AACN consensus conference statement on neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, *35*(6), 1053–1106. https://doi.org/10.1080/13854046.2021.1896036

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science (New York, N.Y.)*, *240*(4857), 1285–1293. https://doi.org/10.1126/SCIENCE.3287615

Taylor, S., Frueh, B. C., & Asmundson, G. J. G. (2007). Detection and management of malingering in people presenting for treatment of posttraumatic stress disorder: Methods, obstacles, and recommendations. *Journal of Anxiety Disorders*, *21*(1), 22–41. https://doi.org/10.1016/j.janxdis.2006.03.016

Thomas, K. M., Hopwood, C. J., Orlando, M. J., Weathers, F. W., & McDevitt-Murphy, M. E. (2012). Detecting feigned PTSD using the Personality Assessment Inventory. *Psychological Injury and Law*, *5*(3-4), 192–201. https://doi.org/10.1007/S12207-011-9111-6/TABLES/3

US Department of Veterans Affairs (2023). *Annual benefits report - Veterans Benefits Administration reports*. https://www.benefits.va.gov/REPORTS/abr/

Wechsler, D. (2009). *Wechsler memory scale-fourth edition*. Pearson.

Whiteside, D. M., & Basso, M. R. (2024). Innovations in performance and symptom validity testing: Introduction to symptom validity sec-

tion of the special issue. *Journal of Clinical and Experimental Neuropsychology*, *46*(2), 81–85. https://doi.org/10.1080/13803395.2024.2346022

Wisco, B. E., Nomamiukor, F. O., Marx, B. P., Krystal, J. H., Southwick, S. M., & Pietrzak, R. H. (2022). Posttraumatic stress disorder in US military veterans: Results from the 2019–2020 National Health and Resilience in Veterans Study. *The Journal of Clinical Psychiatry*, *83*(2), 39779. https://doi.org/10.4088/JCP.20M14029

Wygant, D. B., Sellbom, M., Ben-Porath, Y. S., Stafford, K. P., Freeman, D. B., & Heilbronner, R. L. (2007). The relation between symptom validity testing and MMPI-2 scores as a function of forensic evaluation context. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *22*(4), 489–499. https://doi.org/10.1016/J.ACN.2007.01.027

Xue, C., Ge, Y., Tang, B., Liu, Y., Kang, P., Wang, M., & Zhang, L. (2015). A meta-analysis of risk factors for combat-related PTSD among military personnel and veterans. *PloS One*, *10*(3), e0120270. https://doi.org/10.1371/JOURNAL.PONE.0120270

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, *3*(1), 32–35. https://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3

Young, G. (2015). Malingering in forensic disability-related assessments: Prevalence 15 ± 15. %*Psychological Injury and Law*, *8*(3), 188–199. https://doi.org/10.1007/s12207-015-9232-4

Young, G., & Erdodi, L. (2024). Forensic prediction of malingering from performance validity tests: Review of Leonhard (2023, a, b, c). *Psychological Injury and Law*, *17*(2), 129–138. https://doi.org/10.1007/s12207-024-09504-9