International Journal of Law and Psychiatry 74 (2021) 101669

Contents lists available at ScienceDirect

International Journal of Law and Psychiatry

journal homepage: www.elsevier.com/locate/ijlawpsy

Specificity and item endorsement rates of personality assessment inventory over-reporting scales across ethnic, gender, and diagnostic groups in a forensic inpatient sample

Mary Elizabeth Wood [a,b,*], C. Adam Coffey [a], David M. Glassmire [a]

[a] Patton State Hospital, Patton, CA, USA
[b] Vanderbilt University Medical Center, Nashville, TN, USA

ARTICLE INFO

Keywords:
PAI
Validity indices
Specificity
Overreporting

ABSTRACT

The accurate evaluation of response style, particularly with respect to overreporting, is imperative in forensic settings wherein an external incentive to feign exists. Given the high cost of false positive errors in this context, as well as the associated cost of false negative errors, evaluators need to ensure that overreporting methods are effective with the unique patient populations with whom they work. Complicating this issue is that forensic samples often differ in predictable ways from the normative samples upon which typical psychological assessment instruments were normed. The purpose of the present investigation was to evaluate the specificity of the overreporting indices on the Personality Assessment Inventory, one of the most commonly used personality inventories, in a forensic sample with no ostensible incentive to feign. Although item endorsement and configural elevations on the Negative Impression Management (NIM) scale and the Malingering Index (MAL) were associated with genuine psychopathology, results indicated that the overall specificity estimates across groups were generally adequate. Further, and consistent with prior research, Rogers Discriminant Function (RDF) performed poorly in this sample.

1. Introduction

The accurate evaluation of response style, particularly with respect to overreporting[1] indices, is important in criminal forensic evaluative settings wherein individuals often have a known external incentive to feign. Estimates of overreporting in this population have varied widely, and often depend on the source of the estimate (i.e., clinician estimates range from 12 to 19% versus 10-–to 25% in research settings; Rogers, Sewell, & Goldstein, 1994; Gothard, Rogers, & Sewell, 1995; Mittenberg, Patton, Canyock, & Condit, 2002; Heinze, 2003; Rogers, Ustad, & Salekin, 1998). Given the estimated prevalence of overreporting, as well as the consequences of an inaccurate determination by the court, it is imperative that clinicians are able to evaluate an individual's response style effectively in order to assist the factfinder in forensic assessments.

The Personality Assessment Inventory, 2nd Edition (PAI; Morey, 2007) is a multiscale inventory of personality and psychopathology often used in clinical and forensic settings. Indeed, the PAI is the second most frequently used multiscale inventory in forensic evaluations (Archer, Buffington-Vollum, Stredny, & Handel, 2006; Neal & Grisso, 2014), second only to the Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989; Butcher, Graham, Ben-Porath, Tellegen, Dahlstrom, & Kaemmer, 2001). The PAI contains several validity scales and indices to detect overreporting, including the Negative Impression Management (NIM) scale, Malingering Index (MAL), and Rogers Discriminant Function (RDF; Rogers, Sewell, Morey, & Ustad, 1996). NIM has some similarities to the MMPI-2 *F* family of scales in that the items were selected based on infrequent endorsement in both normative and clinical samples (Morey, 2007). The PAI manual explicitly notes that although NIM can be considered an indicator of overreporting, scores tend to be differentially elevated in populations with severe psychopathology and, therefore, may be elevated due to symptoms of a mental disorder. This is of great

* Corresponding author at: 1601 23rd Ave. S., 3rd Floor, Nashville, TN 37212, USA.
E-mail address: mary.e.wood@vumc.org (M.E. Wood).

[1] In this paper, the terms overreporting and overreporters or feigning and feigners will be used as opposed to malingering and malingerers, for the sole fact that psychological assessment instruments are unable to extricate an individual's motivation and, therefore, only assess overreporting or feigning (see Rogers & Bender, 2012).

https://doi.org/10.1016/j.ijlp.2020.101669
Received 26 May 2020; Received in revised form 1 December 2020; Accepted 7 December 2020
Available online 17 December 2020

concern when considering its use in forensic inpatient settings wherein most patients have been diagnosed with a schizophrenia spectrum illness (e.g., Hoge et al., 1997; Melton, Petrilla, Poythress, & Slobogin, 2007; Viljoen & Zapf, 2002), potentially resulting in artificial NIM elevations reflective of severe psychopathology rather than overreporting.

The MAL Index is calculated based on eight profile configurations associated with feigning (Morey, 2007) and was developed to provide an indicator of overreporting independent of genuine psychiatric illnesses. The relative *T* score differences upon which MAL is based are purported to reflect simulations of severe mental disorders, as differences this large are unlikely to occur in genuinely mentally ill individuals. RDF (Rogers et al., 1996) is a regression-based indicator of overreporting that was derived following a contrast between sophisticated instructed overreporters (i.e., doctoral students with training in advanced psychopathology) attempting to simulate mental illness (Rogers et al., 1996). Although RDF has produced consistently large effect sizes in simulation studies, its utility in criterion group validation studies has been limited, leading Morey (2007) and other authors (including Rogers et al., 1996) to caution against its use in clinical forensic settings.

1.1. Research with the PAI overreporting indices

Research generally supports the utility of the PAI overreporting scales/indices. In a meta-analysis of 26 studies conducted between 1993 and 2008, Hawes and Boccaccini (2009) reported that NIM yielded a high classification rate of 0.79 when using a cut-score of ≥81 *T*, with sensitivity and specificity estimates of 0.73 and 0.83, respectively. Relatedly, a MAL Index score of ≥3 yielded a classification rate of 0.71, with sensitivity and specificity estimates of 0.58 and 0.86, respectively. Beyond this, MAL Index scores of 4 and 5 yielded specificity estimates of 0.99 and 1.0. Hawes and Boccaccini (2009) found that RDF was no better than chance in criterion group studies. Therefore, they concluded that clinicians should avoid using RDF in clinical practice until additional research on the functioning of the RDF in criterion groups studies was available. More broadly, the authors observed smaller effect sizes across all three validity measures in criterion groups studies relative to simulation studies, illustrating the need for greater understanding of how these scales perform in real-world evaluative contexts. In existing studies with clinical samples, the NIM and MAL scales produced significantly larger effects for detecting overreporting of more severe disorders (i.e., psychosis) relative to mood or anxiety-related disorders, while RDF produced smaller effects for each group that were not significantly different from one another (Hawes & Boccaccini, 2009).

Several studies included in Hawes and Boccaccini's (2009) meta-analysis warrant separate discussion given that they were conducted with individuals who presented in forensic evaluative contexts. First, for example, in one sample of 154 individuals referred for federal pre-trial court ordered evaluations, Boccaccini, Murrie, and Duncan (2006) found that an NIM cutoff of ≥81 *T* resulted in a sensitivity of 0.91 and a specificity of 0.72 in identifying overreporting on the Structured Interview of Reported Symptoms (SIRS; Rogers, Bagby, & Dickens, 1992), a well-validated measure of overreporting. Although not as effective as NIM, the MAL Index performed well, with sensitivity and specificity estimates of 0.89 and 0.61, respectively, for a score of ≥2. Higher scores on the MAL index (i.e., ≥ 5) resulted in very few false positive errors ($n = 3$). In contrast to the impressive performance of NIM and MAL, RDF performed no better than chance, leading the authors to conclude that this measure is not appropriate in forensic-correctional settings.

Similar support was found for NIM with a sample of 116 criminal defendants from the federal prison system (Kucharski, Toomey, Fila, & Duncan, 2007) who were divided into overreporting and honest responding groups based on SIRS scores. Both NIM and MAL produced large effect sizes, whereas RDF failed to reach statistical (or practical) significance. Similarly, NIM was the only scale to produce acceptable classification accuracy. An NIM cutoff of ≥84 *T* resulted in sensitivity of 0.84 and specificity of 0.82. Kucharski et al. (2007) therefore opined that NIM was appropriate for identifying feigning, though neither MAL nor RDF were endorsed for this purpose.

Together, this body of literature provides evidence for the utility of NIM in clinical settings, partial support for the MAL Index (particularly at higher thresholds), and minimal to no support for RDF. One issue complicating the interpretation of previous research, however, is that individuals who present for criminal forensic evaluations often differ in important ways from the normative samples on which these measures were developed and the research samples on which they were subsequently studied. These differences range from demographic differences on variables like ethnicity, education, and age, to clinical status variables like psychiatric diagnoses and presence of symptoms, to the context of the evaluation. Previous research indicates that the majority of individuals in pre- and post-trial forensic evaluation settings meet the diagnostic criteria for a Schizophrenia Spectrum Disorder (Cooper & Zapf, 2003; Hoge et al., 1997; Nicholson & Kugler, 1991; Viljoen & Zapf, 2002; Warren et al., 2006), and that individuals adjudicated incompetent to stand trial (IST) and individuals acquitted pursuant to not guilty by reason of insanity (NGRI) share demographic and diagnostic characteristics given the known overlap between these two groups (Boehnert, 1989; Melton et al., 2007).

Despite the high prevalence of individuals with schizophrenia spectrum diagnoses in criminal forensic settings, only 5.6% of the clinical sample employed in the development of the PAI had a primary diagnosis of Schizophrenia (Morey, 2007). Additionally, most studies investigating the PAI validity scales obtained specificity estimates from nonclinical subjects who were instructed to respond honestly. The lack of research on the specificity of overreporting indices in forensic psychiatric settings can lead to the adoption of cutoff scores on these indices that have higher false positive rates due to differences in endorsement rates of test items across forensic and non-forensic settings (Glassmire, Jhawar, Burchett, & Tarescavage, 2017). Moreover, although only two previous studies investigated the utility of PAI overreporting measures in samples of *pretrial* federal defendants (Boccaccini et al., 2006; Kucharski et al., 2007), the diagnostic composition of the samples was not provided, the samples were drawn from the same larger population of federal inmates, and both studies relied on the SIRS as the criterion to establish lack of overreporting to derive specificity estimates, as participants in both studies had an ostensible reason to feign. Tarescavage and Glassmire (2016) outlined the limitations of deriving specificity estimates for overreporting measures with pretrial forensic samples given the imperfect nature of measures used to screen out potential feigners from the sample. As such, research is needed that investigates the specificity of PAI overreporting indicators among criminal forensic psychiatric examinees without an ostensible reason to feign.

1.2. Current study

The purpose of the current study was to evaluate the specificity of NIM, the MAL Index, and RDF in a post-adjudication forensic inpatient sample with no known external incentive to feign. These individuals are presumed to have genuine mental illness given that the nature of their legal commitment was predicated on the court's finding of the presence of a mental disorder. Previous studies (e.g., Glassmire, Jhawar, Burchett, & Tarescavage, 2017; Glassmire, Toofanian Ross, Kinney, & Nitch, 2016; Weinborn, Orr, Woods, Conover, & Feix, 2003) have used NGRI and mentally disordered offender (MDO) patients in a similar manner to calculate specificity estimates for other measures, as patients who have been exonerated of their charges due to legal insanity (NGRI) or who are committed to the hospital as part of their parole (MDO) have no known incentive to overreport due to the nature of their hospital commitment (i.e., the need to demonstrate psychiatric stability in order to secure release from the hospital). Beyond this, such individuals are similar to other criminal forensic examinees with regard to important variables such as demographics, criminal histories, diagnoses, and other clinically relevant variables.

It is important to identify specificity estimates in this context to obtain generalizable specificity estimates for various cutoffs. This is particularly imperative given the lack of data on the utility of these measures in individuals with schizophrenia spectrum disorders, as well as the concern that NIM might reflect, to some degree, genuine psychopathology. Given prior research in this area, it was expected that chosen NIM cutoff scores would have higher specificity values than both MAL and RDF cutoff scores in this sample, with minimal to no support for RDF.

We also evaluated whether specificity estimates varied across demographic subgroups (i.e., gender, ethnicity, and diagnostic subgrouping) given the known demographic differences between forensic inpatient samples and the samples used in the development and cross-validation of PAI overreporting indices. In addition to specificity, item endorsement (for NIM items) and configural elevations (for the MAL Index) were evaluated to understand which, if any, elements of these indices contributed to scale elevations across groups. Previous research on the MMPI-2-RF Fp-r scale in forensic settings indicates that clinicians can have increased confidence in overall scale elevations because most items on that scale have low endorsement rates across different ethnic and gender groups among forensically committed individuals with no incentive to feign (Glassmire, Jhawar, Burchett, & Tarescavage, 2017). Therefore, an additional focus of this study was to investigate whether the NIM items from the PAI demonstrate similarly low endorsement rates across ethnicity and gender among forensic psychiatric patients. Finally, it was hypothesized that NIM and MAL would be associated with genuine psychopathology (as measured by PAI clinical scales), consistent with concerns outlined in the PAI manual (Morey, 2007), with little association between RDF and PAI clinical scale elevations.

2. Method

2.1. Participants

Archival PAI data were available for 269 psychiatric inpatients whose legal status was previously adjudicated and whose discharge criteria from the hospital required them to demonstrate psychiatric stability. The two legal commitments used in this study included NGRI, as well as commitments under California's MDO statute. The MDO statute allows for post-prison hospital commitment at the time of parole because they are considered to be dangerous as a result of a severe mental disorder. Release from the hospital for NGRI and MDO patients is predicated on the patient demonstrating that he or she no longer represents a danger to others by reason of a severe mental disorder. The PAI was administered as part of routine clinical or forensic assessment during their commitment. Given that the key assumption of the current study's research design is predicated on individuals having a genuine mental illness that they are portraying honestly, data for individuals who produced inconsistent PAI protocols were identified and excluded. Specifically, individuals with an inconsistency score (INC) greater than or equal to 73 T ($n = 29$), an infrequency score (INF) greater than or equal to 75 T ($n = 45$), and/or more than 17 unanswered questions on the PAI ($n = 3$); these three criteria were taken from the instrument's manual outlining the detection of invalid protocols due to inconsistency, irrelevant responding, lack of comprehension, or incompleteness (Morey, 2007), and have been used to screen out inconsistent protocols in previous research (e.g., Boccaccini et al., 2006; Kucharski et al., 2007). It is important to note that these indicators reflect seemingly careless or random responding rather than content-based distortions (i.e., systematic over- or underreporting).

Additionally, data were excluded from 20 individuals who produced positive impression management scores (PIM) of greater than or equal to 68 T, which the PAI professional manual recommends as the cutoff for identifying individuals who attempt to portray themselves as exceptionally free of common shortcomings. Elevations on PIM can reflect a lack of insight, commensurate with anosognosia, and/or deliberate attempts to minimize one's difficulties. These individuals were excluded due to concerns that their demonstrated response style would decrease the rate of false positive classifications in the overreporting scales under study. Lastly, data were excluded from participants who had a *DSM-IV-TR* (American Psychiatric Association, 2000) designation of Malingering listed on their diagnostic formulation ($n = 3$) on the date of testing.

Examination of potential differences between included and excluded participants was conducted based on the source of exclusion. Participants who were excluded due to evidence of random or careless responding (i.e., response bias independent of item content, evidenced by elevations on INC/INF or excessive unresponsiveness; $n = 77$) did not differ from included participants based on gender or age, but there was an association with race/ethnicity ($\chi^2(df = 1) = 14.80$, $p < .001$, Cramer's V = 0.24). Indeed, non-white participants were 3.02 times more likely to be excluded than white participants due to evidence of inconsistent or careless responding. In contrast, female participants were 4.16 times more likely to be excluded than male participants due to underreporting ($\chi^2(df = 1) = 10.14$, $p < .001$, Cramer's V = − 0.19). There was not a significant association between underreporting and race/ethnicity or age.

The final sample included data from 169 PAI profiles. The final sample was 66.3% male ($n = 112$) and 33.7% female ($n = 57$). The average age of patients in the sample was 42.76 years ($SD = 12.07$), with a range of 18 to 74. The self-identified racial and ethnic composition of the sample was 54.4% Caucasian ($n = 92$), 23.7% African American ($n = 40$), 15.4% Hispanic/Latino ($n = 26$), 4.7% Asian American ($n = 8$), and 1.8% Other ($n = 3$). The legal commitment status of the sample was 60.9% NGRI ($n = 103$) and 39.1% MDO ($n = 66$). NGRI and MDO patients did not differ significantly in terms of age or race/ethnicity ($p < .05$), though there was a significant association between commitment status and gender ($\chi^2(df = 1) = 47.95$, $p < .001$, Cramer's V = 0.53, with significantly more male than female insanity acquittees (OR = 0.38).

Patients were also classified according to psychiatric diagnoses of record, provided by interdisciplinary treatment teams consisting of a psychologist, psychiatrist, social worker, rehabilitation therapist, and nursing staff who had access to 24-hour observations of the patients during treatment. Of note, diagnoses were taken from the records from the date of testing (and prior to the test results being scored) to avoid criterion contamination. On average, patients were administered the PAI approximately 3 ¾ years following their admission to the hospital ($M = 1397.07$ days; $SD = 1740.26$ days; Range = 2–8262 days), suggesting that patients were well-known to their interdisciplinary treatment teams at the time that the diagnoses were rendered. The *DSM-IV-TR* (APA, 2000) was in use at the time that diagnoses were rendered.

Most patients had a primary diagnosis reflecting either a Psychotic Disorder ($n = 73$; 43.2%), or a disorder reflecting both mood and psychotic symptoms (e.g., Bipolar Disorder with Psychotic Features or Schizoaffective Disorder; $n = 70$; 41.4%). Smaller numbers of patients were diagnosed with a Mood Disorder alone ($n = 19$; 11.2%), or another diagnosis altogether ($n = 5$; 3%). See Supplemental Table 1 for the breakdown of primary psychiatric diagnoses in the present sample. In addition to examining the breakdown of primary diagnoses, patients were categorized into groups reflecting the totality of assigned diagnoses (i.e., primary, secondary, etc. diagnoses of record). In total, most patients were categorized in either the Mood/Psychotic Disorder Group ($n = 9$; 45.6%) or the Psychotic Disorder Only Group ($n = 71$; 42%). A smaller number of individuals were diagnosed with a Mood Disorder Only ($n = 16$; 9.5%); a total of five patients were not categorized in any of the three categories, as they had no diagnoses reflecting a mood and/or psychotic disorder.

The two former categories (i.e., Mood/Psychotic Disorder and Psychotic Disorder Only) served as the basis for the calculation of item endorsement rates in subsequent analyses, as too few participants were categorized in the Mood Disorder Only group for meaningful analyses with this subgroup (i.e., $n = 14$). There were no significant associations between overall diagnostic category and race/ethnicity ($p = .32$) or

between race/ethnicity and gender ($p = .81$). There was a significant association between overall diagnostic category and gender, ($\chi^2(df = 2) = 9.92$, $p = .02$, Cramer's V = 0.24). Specifically, females were more likely to be diagnosed with Mood and Psychotic/Mood Disorders, whereas men were more likely to be diagnosed with Psychotic Disorders.

In addition to psychiatric diagnoses, 49 participants (29.0%) were diagnosed with a personality disorder, which is consistent with the diagnostic makeup of inpatient and forensic samples (Black et al., 2007; de Ruiter & Trestman, 2006; Warren & South, 2009; Widiger & Weissman, 1991). The majority ($n = 26$; 53.1%) were diagnosed with Antisocial Personality Disorder (ASPD), while about one-third ($n = 15$; 30.6%) were diagnosed with Personality Disorder, Not Otherwise Specified (NOS). Fewer participants had diagnoses of Borderline ($n = 7$), Narcissistic ($n = 1$), Paranoid ($n = 2$), or Schizotypal ($n = 1$) Personality Disorders. Due to the nature of the sample (i.e., patients committed to an inpatient psychiatric hospital due to a qualifying mental disease or defect, most often of a psychiatric etiology), the focus of the present study (i.e., examination of overreporting indices and their relationship to severe psychopathology), and the heterogeneity among personality disorders and the variable rates of representativeness in this sample, analyses were not conducted on this subset of participants.

2.2. Measures

The PAI (Morey, 2007) is a 344-item multiscale inventory of personality and psychopathology that includes four validity scales, along with a variety of clinical, treatment, and interpersonal scales totaling 22 completely independent (i.e., non-overlapping) scales. In addition to the standardization sample, the PAI also includes a large clinical comparison sample ($n = 1246$) derived from a variety of settings including outpatient mental health (34.6%), inpatient mental health (24.9%), and substance abuse treatment programs (15.4%), for example. The psychometric characteristics of NIM, MAL, and RDF, the focus of the present investigation, are adequate. Coefficient alphas for NIM range from 0.63 to 0.74, depending on the sample used (i.e., college versus clinical samples, respectively), with an unweighted mean of 0.60 across various research studies (Morey, 2007). Coefficient Alpha for NIM in the current sample was 0.73. Classification accuracy statistics for NIM, MAL, and RDF from previous studies were presented in the Introduction section.

2.3. Procedure

The current research project was approved by the California Committee for the Protection of Human Subjects, and all ethical guidelines were followed in the collection, maintenance, and analysis of the data. Archival data were exported into SPSS 26.0, a statistical software program, for all subsequent analyses. Specificity values were calculated for multiple cutoff scores on NIM, the MAL Index, and RDF for the overall sample and for various demographic and diagnostic subgroups (i.e., by gender, ethnicity, and diagnostic category). Additionally, endorsement rates were calculated for the items comprising the NIM subscale for each group mentioned above. Item endorsement was conceptualized as any response earning the patient a positive score on the item (i.e., any response option not earning a score of 0). The reason that this was selected (i.e., as opposed to dichotomizing scores of 0 and 1 from scores of 2 and 3) is that any increase in point value corresponds to an elevation on the scale. In addition, individual configuration items from the MAL Index were calculated for the total sample and among the subgroups to evaluate feature elevations within each sample. Finally, bivariate correlations were computed between the three validity indices and the clinical scales to evaluate the association between the scales and psychopathology.

3. Results

3.1. Specificity estimates

Specificity values were calculated for a range of cutoffs on NIM and MAL (see Table 1). The specificity estimates presented in Table 1 are generally consistent with those derived from previous research. The NIM cutoff generally identified as sufficient in previous research (≥ 81 *T*) was associated with a specificity estimate of 0.87 for the total sample. Among the various demographic groups, specificity estimates for NIM ≥ 81 *T* ranged from 0.80 to 0.92, meaning up to a 12% variability in the rate of false positive errors depending on the subgroup. A cutoff score of ≥ 92 *T* (the score generally used to indicate a "cry for help;" Morey, 2007) was associated with specificity estimates ranging from 0.90 to 0.97, with an overall specificity estimate of 0.95. Of interest, raw scores on NIM were highly negatively skewed, with 19% of the sample earning a raw score of 0.

Consistent with prior research, the MAL Index demonstrated high specificity estimates at almost all the cutoff scores examined (i.e., specificity estimates ranging from 0.93 to 0.1.00 for all groups at all cutoffs ≥ 3). Raising the cutoff to ≥ 4 resulted in estimates ranging from 0.95 to 1.00, which is consistent with previous research indicating that scores in this range can increase confidence in determinations of feigning to a very high degree (particularly given specificity estimates hovering close to, or at, 1.00). In addition to high overall specificity rates, there was minimal variability across subgroups, with false positive rates varying by 7% at a cutoff of ≥ 3, and 5% at the cutoff of ≥ 4.

To calculate specificity estimates, RDF values were transformed to *T* scores using the values from the PAI profile form for adults (e.g., an RDF value of 1.25 corresponds to a *T* score of approximately 71). Using the standard interpretative guideline published in the manual (i.e., scores greater than 59 *T* as indicative of overreporting), the specificity of RDF was 0.80. Scores at this cutoff (i.e., 59 *T*) were associated with specificity estimates ranging from 0.66 to 0.90 depending on the subgroup. Increasing the cutoff to 65 *T,* a more typical threshold for interpreting *T* score elevations, resulted in an overall specificity estimate of 0.94. Increasing the cutoff to ≥ 75 *T* (i.e., RDF $\geq$ 1.75) resulted in specificity estimates of 0.99 across all subgroups.

3.2. Item endorsement: NIM

In addition to examining specificity estimates, it was important to identify the specific elements of each overreporting scale to understand which items and/or features contributed to elevations in this the overall sample and subsamples. NIM items were categorized as 'endorsed' for any response earning a score greater than 0. Consistent with the endorsement rates in clinical samples used to develop the *Fp* scale on the MMPI-2 (Arbisi & Ben-Porath, 1995), endorsement rates greater than 20% were considered significant.

Five items were endorsed by greater than 20% of the entire sample (Table 2). Over half the sample endorsed one of the items, with endorsement rates ranging from 54.5% to 67.5% across the various subgroups. Similarly, two items were endorsed by approximately one-third of the sample, with generally consistent endorsement rates across the subgroups. The content of the frequently endorsed items reflected self-reported memory loss, a sudden onset of psychological problems, and the belief that others do not appreciate their suffering. Consistent with the caveat printed in the measure's professional manual (Morey, 2007), these items appeared to reflect normative reports associated with severe psychopathology. Notably, the mood/psychotic disorder subgroup produced endorsement rates of greater than 20% on eight of nine NIM items, and approximately two-thirds (67.5%) this group endorsed one item (Item 169).

Table 1
Specificity estimates for NIM and MAL for the overall sample and demographic groupings.

Cutoff score	Female	Male	African American	Caucasian	Hispanic	Mood & psychotic	Psychotic	Total sample
NIM								
$T \geq 96$	0.98	0.97	0.93	0.99	1.00	0.95	1.00	0.97
$T \geq 92$	0.97	0.94	0.90	0.96	0.96	0.92	0.96	0.95
$T \geq 88$	0.95	0.93	0.90	0.95	0.92	0.92	0.94	0.94
$T \geq 84$	0.91	0.91	0.83	0.94	0.92	0.90	0.92	0.91
$T \geq 81$	0.90	0.89	0.80	0.91	0.92	0.90	0.89	0.87
$T \geq 77$	0.84	0.88	0.78	0.90	0.89	0.87	0.87	0.83
$T \geq 73$	0.84	0.79	0.75	0.87	0.80	0.81	0.86	0.79
$T \geq 65$	0.75	0.73	0.68	0.80	0.65	0.68	0.78	0.74
MAL index								
≥7	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
≥6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
≥5	1.00	1.00	0.98	1.00	1.00	0.99	1.00	0.99
≥4	0.99	0.98	0.95	1.00	1.00	0.97	1.00	0.99
≥3	0.96	0.98	0.93	0.98	1.00	0.94	1.00	0.97
RDF								
$T \geq 85$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$T \geq 75$	1.00	0.99	1.00	1.00	1.00	1.00	1.00	0.99
$T \geq 73$	1.00	0.99	0.98	1.00	1.00	1.00	1.00	0.99
$T \geq 65$	0.93	0.95	0.98	0.96	0.85	0.96	0.92	0.94
$T \geq 59$	0.77	0.80	0.83	0.79	0.77	0.90	0.66	0.80

Note. Female ($n = 57$). Male ($n = 112$). African American ($n = 43$). Caucasian ($n = 102$). Hispanic ($n = 27$). Mood & Psychotic ($n = 91$). Psychotic ($n = 75$). Total Sample ($n = 188$).

Table 2
NIM item endorsement by demographic groups.

Item	Female	Male	African American	Caucasian	Hispanic	Mood & psychotic	Psychotic	Total sample
9	14.1	12.5	**20.0**	13.0	3.8	**22.1**	5.6	13.1
49	8.8	16.9	17.5	8.7	**23.1**	19.2	11.3	14.2
89	**22.8**	15.2	15.0	17.4	19.2	**23.4**	15.5	17.8
129	19.3	**25.0**	**22.5**	**21.7**	19.2	**27.3**	**21.1**	**23.1**
169	**57.9**	**58.9**	**62.5**	**56.5**	**57.7**	**67.5**	**53.5**	**58.0**
209	**36.9**	**41.1**	**41.5**	**34.8**	**38.5**	**45.5**	**36.6**	**39.0**
249	15.8	15.2	**25.0**	12.0	15.4	**23.4**	7.0	15.4
289	**35.1**	**23.2**	**35.0**	**22.8**	**34.6**	**26.0**	**32.4**	**27.2**
329	**42.1**	**39.3**	**42.5**	**41.3**	**26.9**	**55.5**	**28.2**	**40.2**

Note. Female ($n = 57$). Male ($n = 112$). African American ($n = 43$). Caucasian ($n = 102$). Hispanic ($n = 27$). Mood & Psychotic ($n = 77$). Psychotic ($n = 71$). Total Sample ($n = 169$). Significant elevations (in **bold**) were operationally defined as endorsement rates greater than 20%.

3.3. Configuration elevations: MAL index

The standard cutoff score for the *MAL Index* (i.e., 3 or greater) was associated with specificity of 0.97 in the overall sample. Like NIM, most patients (45%) earned a raw score of 0 on the MAL Index. Approximately one-third (31.4%) earned a raw score of 1, and 14.8% earned a raw score of 2. A significant percentage of the overall sample elevated Feature 7 (28.2%) of the MAL Index (Table 3), which reflects a relative difference in *PAR* (paranoia) subscale scores greater than 15 *T* (i.e., a higher score on the subscale designed to measure persecutory ideation relative to resentment). Notably, higher endorsement rates were observed in male (35.6%), African American (37.2%), Hispanic (40.7%) and combined mood/psychotic disorder (33%) subgroups. Additionally, a significant percentage of the overall sample elevated Feature 5 (27.8%) of the MAL index, which reflects a relative difference in *MAN* (mania) subscale scores greater than 15 *T* (i.e., a higher score on the subscale designed to measure related to strained relationships due to irritability and frustration relative to grandiose self-appraisal). Subgroup analysis was generally consistent with the observed increased proportion of elevations on Feature 7, with male (36.6%), African American (35%), Hispanic (42.3%), and combined mood/psychotic disorder (31.2%) groups demonstrating greater percentages of individuals with elevations.

Table 3
MAL index frequency by demographic groups (% present).

Feature	Female	Male	African American	Caucasian	Hispanic	Mood & psychotic	Psychotic	Total sample
1	1.8	2.7	**7.5**	1.1	0	5.2	0	2.4
2	12.3	16.1	**25.0**	12.5	15.4	18.2	11.3	14.8
3	12.3	12.5	15.0	6.5	**26.9**	9.1	19.7	12.4
4	17.0	17.0	15.0	15.2	15.4	**22.1**	9.9	16.6
5	10.5	**36.6**	**35.0**	**20.7**	**42.3**	**31.2**	**28.2**	**27.8**
6	10.5	8.9	10.0	13.0	0	11.7	7.0	9.5
7	15.7	**35.6**	**37.2**	**21.6**	**40.7**	**33.0**	**26.7**	**28.2**
8	7.0	9.8	7.5	**37.5**	**23.1**	10.4	7.0	8.9

Note. Female ($n = 57$). Male ($n = 112$). African American ($n = 40$). Caucasian ($n = 92$). Hispanic ($n = 26$). Mood & Psychotic ($n = 77$). Psychotic ($n = 71$). Total Sample ($n = 169$). Feature 1 (*INF* minus *ICN* $\geq$ 15 *T*). Feature 2 (*NIM* $\geq$ 110 *T*). Feature 3 (*NIM* minus *INF* $\geq$ 20 *T*). Feature 4 (*DEP* $\geq$ 85 *T* and *RXR* $\geq$ 45 *T*). Feature 5 (*MAN-I* minus *MAN-G* $\geq$ 15 *T*). Feature 6 (*PAR-P* minus *PAR-H* $\geq$ 15 *T*). Feature 7 (*PAR-P* minus *PAR-R* $\geq$ 15 *T*). Feature 8 (*ANT-E* minus *ANT-A* $\geq$ 10 *T*).

3.4. Association between clinical scales and overreporting indices

Table 4 displays correlations between clinical scale elevations and the over-reporting indices (i.e., NIM, MAL, and RDF). Examination of the values reveals strong, positive correlations between all the clinical scales and NIM and all but one clinical scale (ALC, $r = 0.19$) and MAL, suggesting a positive association between endorsement of psychopathology and endorsement of items presumed to be indicative of over-reporting. Three of the clinical scales were significantly correlated with RDF at the $p < .05$ level.

4. Discussion

The present study was designed to investigate the utility of, and specificity estimates for, the overreporting indices on the PAI in a demographically diverse forensic inpatient sample with no known or obvious incentive to overreport. Because the present sample included only individuals who had no known incentive to overreport and was screened to exclude individuals with indications of possible over-reporting, underreporting, and/or inconsistency, elevations on PAI overreporting scales are likely to be reflective of false positive errors in identifying overreporting. Indeed, the biggest strength of the present study was the use of a culturally and diagnostically diverse real-world clinical-forensic sample, suggesting greater generalizability with respect to forensic inpatient populations.

Five of the nine NIM items were endorsed by over 20% of the sample (the threshold at which infrequent endorsement is generally identified for items on similar overreporting scales from the MMPI-2-RF; Ben-Porath, 2012), and this was consistent across demographic groups. The content of these items reflected general distress and correlates of psychosis (e.g., acute memory loss, abrupt onset of symptoms), which is unsurprising given the diagnostic composition of forensic samples. Further, this finding is consistent with the caveat in the instrument's manual, that elevations may reflect severe psychopathology.

Despite high rates of NIM item endorsement, the specificity of the overall scale was generally adequate in the current sample. In the present study, endorsement was defined as any response option earning a score of '1' or more, meaning that only responses of '0' ("False, Not at All True") were conceptualized/coded as non-endorsements. This discrepancy explains the relatively high rate of individual item endorsements, coupled with generally acceptable specificity estimates for the full scale. Looking at the five items that were endorsed at a rate greater than 20% (i.e., items 129, 169, 209, 289, and 329), endorsement rates would shift to 10.0% 34.9%, 10.1%, 8.9%, and 17.2% if only response options '2' ("Mainly True") and '3' ("Very True") were operationally defined as endorsement. In other words, in three of the five instances, a large proportion of the individuals who 'endorsed' the item selected response option '1' ("Slightly True"). This finding indicates that the multi-level Likert response options on the PAI may add to the specificity of NIM by increasing the overall variability of raw scores, thus providing more opportunity to distinguish between forensic examinees who are responding honestly versus overreporting.

Despite generally adequate specificity estimates for NIM with the overall sample, these values varied across demographic subgroups, with relative reductions in specificity for certain subgroups (i.e., a 25% false positive error rate for the African American subsample). Increasing the cutoff reduced the false positive error rates across the various subgroups, but even at a cutoff of ≥96 T, the Mood & Psychotic and African American samples had specificity estimates of 0.95 and 93. respectively. Further, it is important to consider the relationship between sensitivity and specificity, as higher cutoffs may result in poor sensitivity. For example, in a sample of federal pretrial inmates, Kucharski et al. (2007) reported a sensitivity estimate of 0.71 and Boccaccini et al. (2006) reported a sensitivity estimate of 0.78 at a cutoff score of ≥92 T. At a score of ≥88 T, Kucharski et al., (2007) reported a sensitivity estimate of 0.81. In other words, raising the cutoffs to these levels increases the possibility of missing up to 20–30% of overreporters.

Consistent with previous research, high specificity values were found for the MAL index at all levels (i.e., from ≥3 and above). Of interest, there were two features on the MAL Index that appeared to be differentially elevated in this sample, namely those that reflected relative differences in elevations on *PAR* and *MAL* subscales. Despite this anomaly, the overall MAL Index demonstrated adequate specificity in the current sample.

In line with a priori predictions, the specificity estimates of RDF at the recommended cutoff (i.e., 59 *T*) were unacceptably low, ranging from 0.66 to 0.90, with an overall estimate of 0.80. The cutoff score would need to be raised to ≥73 *T* to achieve acceptable specificity. Notably, the specificity of the RDF at the recommended cutoff of ≥59 *T* was poor in the psychotic disorder subgroup (0.66), but in the adequate range (0.90) for the mood/psychotic group. Further, the specificity of the RDF improved to 0.96 when the cutoff was raised to 65 *T*. These findings suggest that use of this particular metric may be more effective in detecting overreporting in the most severe presentations of psychotic-spectrum illness (i.e., schizoaffective disorder) relative to purely psychotic disorders (i.e., schizophrenia), although replication of these findings is necessary in additional inpatient settings.

In the current study, both NIM and MAL significantly correlated with all the clinical scales. This finding is consistent with the caveat outlined in the PAI manual (Morey, 2007), that NIM elevations may reflect genuine psychopathology is some cases. Although MAL was designed to be less directly associated with genuine psychopathology, the present results suggest MAL elevations were significantly associated with elevations on PAI clinical scales. As expected, RDF was least associated with the clinical scales, reflecting the stated intention for which the index was designed.

Table 4
Correlations of NIM and MAL index with clinical scales.

Clinical scale	*NIM*		*MAL*		*RDF*	
	r	*p*	*r*	*p*	*r*	*p*
SOM	0.60	<0.001	0.42	<0.001	−0.08	0.29
ANX	0.66	<0.001	0.42	<0.001	0.05	0.48
ARD	0.65	<0.001	0.45	<0.001	−0.04	0.57
DEP	0.63	<0.001	0.37	<0.001	0.25	<0.001
MAN	0.59	<0.001	0.51	<0.001	−0.18	0.02
PAR	0.54	<0.001	0.47	<0.001	0.25	<0.001
SCZ	0.71	<0.001	0.52	<0.001	0.06	0.41
BOR	0.69	<0.001	0.41	<0.001	−0.05	0.47
ANT	0.60	<0.001	0.47	<0.001	−0.07	0.37
ALC	0.30	<0.001	0.19	0.002	0.04	0.63
DRG	0.41	<0.001	0.24	<0.001	0.01	0.99

Note. SOM = Somatic Complaints; ANX = Anxiety; ARD = Anxiety-Related Disorders; DEP = Depression; MAN = Mania; PAR = Paranoia; SCZ = Schizophrenia; BOR = Borderline Features; ANT = Antisocial Features; ALC = Alcohol Problems; DRG = Drug Problems.

4.1. Limitations & future directions

The present study has some limitations that should be considered when interpreting the results. First, due to the archival nature of the study and the use of consecutively administered PAI profiles, there was no way to determine the representativeness of the sample within the larger population of individuals hospitalized at the facility. Forensic patients who were administered the PAI during their hospitalization may represent a higher functioning group of patients, at least to the degree that they possessed the requisite cognitive skills and psychiatric stability to complete the PAI. However, because the initial sample included all individuals who were committed as NGRI or MDO and subsequently administered the PAI during their hospitalization, the sample is representative of the types of patients who are more likely to be administered multi-scale measures such as the PAI in inpatient

forensic settings.

In addition, it is possible that some individuals in this sample engaged in some level of response distortion. Although this sample was selected because there was no identifiable external incentive to over-report and protocols were removed if there was any suspicion of over-reporting noted by the treatment team, it is possible that some individuals were not effectively screened out. However, if any over-reporters were included in the sample, the effect would have been an artificial inflation of false positive rates in this sample, thereby lowering the specificity estimates and resulting in conservative estimates. In contrast, and perhaps the more salient issue, is the possibility that some participants engaged in the opposite style of response distortion – underreporting. Twenty participants were excluded based on PIM scores above the manual-identified cutoff for invalidity, though 51 participants fell in the range reflective of moderate elevations on this scale. In contrast to the impact of including overreporters, participants who systematically underreported would result in inflated estimates of specificity, thereby producing artificially lower false positive rates. It is presently unknown if there were participants in this sample who engaged in this type of response bias, and to what extent, though future research should seek to address this issue, ideally by replicating and extending the current study's findings by comparing PAI overreporting indices to well-established measures of feigning (i.e., SIRS or SIRS-2; Rogers et al., 1992, Rogers, Sewell, & Gillard, 2010) in forensic inpatient samples.

Finally, it is notable that several subgroups included in our analyses contained a relatively small numbers of individuals (i.e., $n = 27$ Hispanic participants). Given the differences observed among these smaller subgroups with respect to specificity rates and item endorsement, additional examinations of PAI overreporting indices that incorporate larger subgroups of individuals from a range of ethnic groups is warranted to further understand how they perform in forensic, culturally diverse populations. In addition, to may be worthwhile for future research to examine these overreporting indices in more diagnostically diverse groups, including larger samples of individuals diagnosed with personality disorders.

5. Conclusions

Despite these limitations, there are important conclusions that can be drawn from this study. First, these results suggest generally adequate specificity values of NIM among forensic inpatients, particularly at higher cutoffs (i.e., $\geq$ 88 T). That said, endorsement of specific items appeared to reflect, to some degree, genuine psychopathology. Consistent with this, NIM was significantly correlated with all the clinical scales, further bolstering the notion that the scale is associated with genuine psychopathology. Evaluators are encouraged to interpret NIM elevations cautiously, with an emphasis on the content of items that contribute to NIM elevations. In contrast, the MAL Index demonstrated impressive specificity values across all demographic subgroupings. Finally, RDF had unacceptably high false positive rates in this sample, supporting recommendations from previous research that this index not be used in forensic settings. Future research should continue to investigate the utility of these validity indices in forensic samples, potentially with a more focused investigation of item elevations on NIM. In addition, it is recommended that future research be conducted using a similar methodology with the addition of independent, stand-alone feigning measures, as well as more structured diagnostic methods to ensure the representativeness of the demographic subgroups. Regardless, results of the present investigation provide support for the utility of both NIM and MAL in forensic inpatient samples.

Author Note

The findings and conclusions in this article are those of the authors and do not necessarily represent the view or opinions of the California Department of State Hospitals or the California Health and Human Services Agency. David Glassmire receives grant funding on unrelated projects from the University of Minnesota Press, publisher of the MMPI-2 and MMPI-2-RF instruments.

Funding

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

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijlp.2020.101669.

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR*. Washington, DC: Author.

Arbisi, P. A., & Ben-Porath, Y. S. (1995). An MMPI-2 infrequent response scale for use with psychopathological populations: The infrequency-psychopathology scale, F(p). *Psychological Assessment, 7*, 424–431. https://doi.org/10.1037/1040-3590.7.4.424.

Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V., & Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of Personality Assessment, 87*, 84–94. https://doi.org/10.1207/s15327752jpa8701_07.

Ben-Porath, Y. S. (2012). *Interpreting the MMPI-2-RF*. Minneapolis, MN: University of Minnesota Press.

Black, D. W., Gunter, T., Allen, J., Blum, N., Armdt, S., Wenman, G., & Sieleni, B. (2007). Borderline personality disorder in male and female offended newly committed to prison. *Comprehensive Psychiatry, 48*(5), 400–405.

Boccaccini, M. T., Murrie, D. C., & Duncan, S. A. (2006). Screening for malingering in a criminal-forensic sample with the personality assessment inventory. *Psychological Assessment, 18*, 415–423. https://doi.org/10.1037/1040-3590.18.4.415.

Boehnert, C. E. (1989). Characteristics of successful and unsuccessful insanity pleas. *Law and Human Behavior, 13*, 13–39. https://doi.org/10.1007/BF01056161.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *Minnesota multiphasic personality inventory—2: Manual for administration and scoring*. Minneapolis: University of Minnesota Press.

Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W. G., & Kaemmer, B. (2001). *MMPI-2 (Minnesota Multiphasic Personality Inventory-2): Manual for administration, scoring, and interpretation*. Minneapolis: University of Minnesota Press. revised edition.

Cooper, V. G., & Zapf, P. A. (2003). Predictor variables in competency to stand trial decisions. *Law and Human Behavior, 27*, 423–436.

Glassmire, D. M., Jhawar, A., Burchett, D., & Tarescavage, A. M. (2017). Evaluating item endorsement rates for the MMPI-2-RF F-r and Fp-r scales across ethnic, gender, and diagnostic groups with a forensic inpatient sample. *Psychological Assessment, 29*, 500–508. https://doi.org/10.1037/pas0000366.

Glassmire, D. M., Toofanian Ross, P., Kinney, K. I., & Nitch, S. R. (2016). Derivation and cross-validation of cut-off scores for patients with schizophrenia spectrum disorders on WAIS-IV digit-span-based performance validity measures. *Assessment, 23*, 292–306. https://doi.org/10.1177/1073191115587551.

Gothard, S., Rogers, R., & Sewell, K. W. (1995). Feigning incompetency to stand trial: An investigation of the Georgia court competency test. *Law and Human Behavior, 19*(4), 363–373. https://doi.org/10.1007/BF01499137.

Hawes, S. W., & Boccaccini, M. T. (2009). Detection of overreporting of psychopathology on the personality assessment inventory: A meta-analytic review. *Psychological Assessment, 21*, 112–124. https://doi.org/10.1037/a0015036.

Heinze, M. C. (2003). Developing sensitivity to distortion: Utility of psychological tests in differentiating malingering and psychopathology in criminal defendants. *Journal of Forensic Psychiatry and Psychology, 14*, 151–177.

Hoge, S. K., Bonnie, R. J., Poythress, N., Monahan, J., Eisenberg, M., & Feucht-Haviar, T. (1997). The MacArthur adjudicative competence study: Development and validation of a research instrument. *Law and Human Behavior, 21*, 141–179. https://doi.org/10.1023/A:1024826312495.

Kucharski, L. T., Toomey, J. P., Fila, K., & Duncan, S. (2007). Detection of malingering of psychiatric disorder with the personality assessment inventory: An investigation of criminal defendants. *Journal of Personality Assessment, 88*, 25–32.

Melton, G. B., Petrilla, J., Poythress, N. G., & Slobogin, C. (2007). *Psychological evaluations for the courts: A handbook for mental health professionals and lawyers* (3rd ed.). New York, NY: Guilford.

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

Morey, L. C. (2007). *Personality assessment inventory (PAI) professional manual* (2nd ed.). Lutz, FL: Psychological Assessment Resources, Inc.

Neal, T. M., & Grisso, T. (2014). Assessment practices and expert judgment methods in forensic psychology and psychiatry: An international snapshot. *Criminal Justice and Behavior, 41*(12), 1406–1421.

Nicholson, R. A., & Kugler, K. E. (1991). Competent and incompetent criminal defendants: A quantitative review of comparative research. *Psychological Bulletin, 109*, 355–370.

Rogers, R., Bagby, R. M., & Dickens, S. E. (1992). *Structured Interview of Reported Symptoms (SIRS) and professional manual*. Odessa, FL: Psychological Assessment Resources.

Rogers, R., & Bender, S. D. (2012). Evaluation of malingering and deception. In I. B. Weiner, & R. K. Otto (Eds.) (2nd ed.,, *Vol. 11. Handbook of psychology: Forensic psychology* (pp. 517–540). New York, NY: Wiley.

Rogers, R., Sewell, K. W., & Gillard, N. D. (2010). *Structured Interview of Reported Symptoms-2 (SIRS-2) and professional manual*. Lutz, FL: Psychological Assessment Resources.

Rogers, R., Sewell, K. W., & Goldstein, A. (1994). Explanatory models of malingering: A prototypical analysis. *Law and Human Behavior, 18*, 543–552.

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*, 629–640.

Rogers, R., Ustad, K. L., & Salekin, R. T. (1998). Convergent validity of the personality assessment inventory: A study of emergency referrals in a correctional setting. *Assessment, 5*, 3–12.

de Ruiter, C. D., & Trestman, R. L. (2006). Prevalence and treatment of personality disorders in Dutch forensic mental health services. *Journal of the American Academy of Psychiatry and the Law, 35*, 92–97.

Tarescavage, A. M., & Glassmire, D. M. (2016). Differences between SIRS and SIRS-2 sensitivity estimates among forensic inpatients: A criterion groups comparison. *Law and Human Behavior, 40*, 488–502.

Viljoen, J. L., & Zapf, P. (2002). Fitness to stand trial evaluations: A comparison of referred and non-referred defendants. *The International Journal of Forensic Mental Health, 1*, 127–138. https://doi.org/10.1080/149999013.2002.10471168.

Warren, J. I., Murrie, D. C., Stejskal, W., Colwell, L. H., Morris, J., Chauhan, P., & Dietz, P. (2006). Opinion formation in evaluating the adjudicative competence and restorability of criminal defendants: A review of 8,000 evaluations. *Behavioral Sciences & the Law, 24*, 113–132. https://doi.org/10.1002/bsl.699.

Warren, J. I., & South, S. C. (2009). A symptom level examination of the relationship between cluster B personality disorders and patterns of criminality and violence in women. *International Journal of Law and Psychiatry, 32*, 10–17.

Weinborn, M., Orr, T., Woods, S. P., Conover, E., & Feix, J. (2003). A validation of the test of memory malingering in a forensic psychiatric setting. *Journal of Clinical and Experimental Neuropsychology, 25*, 979–990. https://doi.org/10.1076/jcen.25.7.979.16481.

Widiger, T. A., & Weissman, M. M. (1991). Epidemiology of borderline personality disorder. *Hospital & Community Psychiatry, 42*, 1015–1021.