OXFORD JOURNALS
OXFORD UNIVERSITY PRESS

Archives of Clinical Neuropsychology 25 (2010) 118–125

Archives of CLINICAL NEUROPSYCHOLOGY

Diagnostic Utility of the Structured Inventory of Malingered Symptomatology to Detect Malingering in a Forensic Sample

Nick M. Wisdom[1], Jennifer L. Callahan[2,*], Terry G. Shaw[3]

[1]*Department of Psychology, Oklahoma State University, Stillwater, OK, USA*
[2]*Department of Psychology, University of North Texas, Denton, TX, USA*
[3]*Private Practice, Tulsa, OK, USA*

*Corresponding author at: Department of Psychology, University of North Texas, 1155 Union Circle no. 311280, Denton, TX 76203, USA.
Tel.: +1-940-565-2671; fax: +1-940-565-4682.
E-mail address: jennifer.callahan@unt.edu (J. L. Callahan).

Accepted 24 December 2009

Abstract

The Structured Inventory of Malingered Symptomatology (SIMS) is a 75-item, self-report measure to be used with individuals at least 18 years of age, which may be utilized to assess potential malingering of psychosis, neurologic impairment, amnesia, low intelligence, or affective disorder. However, no studies in the literature have examined the diagnostic validity of this instrument in a known-groups design involving medicolegal referrals diagnosed as malingering using the criteria outlined by Slick and colleagues [Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, *13*, 545–561]. The current known-groups, archival study was designed to examine the effectiveness of the SIMS at identifying malingering in patients involved in personal injury lawsuits or disability claims. Findings indicate that a higher cut score on the SIMS for identification of malingering may be appropriate for use with this population. Limitations, clinical implications, and suggestions for further research are discussed.

Keywords: Malingering; Symptom validity testing; Forensic neuropsychology; Assessment

Introduction

The value and utility of neuropsychological examinations are inherently dependent on a patient's ability and willingness to maintain adequate effort and appropriate intent during the assessment. Unfortunately, the temptation to underperform or exaggerate is substantially influenced by the presence of external incentives (Binder & Rohling, 1996), be it monetary compensation, avoidance of criminal prosecution, or avoidance of military duty (Rogers, 1997). For this reason, the assessment of response validity is an essential component of a medically necessary evaluation (Bush et al., 2005) and there has now been a proliferation of research dedicated to the detection of symptom exaggeration, fabrication, and/or insufficient effort (Heilbronner et al., 2009; Sweet, King, Malina, Bergman, & Simmons, 2002). It is important to remember, however, that less than adequate effort exists on a continuum ranging from fatigue-related declines in performance all the way to outright malingering. Although psychiatric definitions of malingering have been available at least since the publication of the DSM III-R (American Psychiatric Association, 1987), it was not until Slick and colleagues (1999) developed a "gold standard" for diagnosing Malingered Neuropsychological Deficit (MND) that neuropsychologists were provided with a more objective means of deciphering genuine versus disingenuous behavior. This was a particularly important step because prevalence estimates of malingering in litigious settings appear to be quite high. Larrabee (2003) synthesized much of this literature and estimated the base rate of malingering to be 40% in cases where there was the potential for secondary gain. Likewise, a survey of the American Board of Clinical Neuropsychology (ABCN) members found a comparable base rate of malingering (Mittenberg, Patton, Canyock, & Condit, 2002).

doi:10.1093/arclin/acp110 Advance Access publication on 28 January 2010

In response to these prevalence estimates, a number of symptom validity tests (SVT) addressing a variety of functions have now been made available to neuropsychologists for use during evaluations (for a review, see Boone, 2007 or Iverson & Binder, 2000). Yet, most of these tend to measure effort in a single domain of function. This is potentially problematic in which honest responders (HR) as well as those who intend to feign impairments are apt to display cognitive, physical, and/or emotional deficits (Holsinger et al., 2002; Nampiaparampil, 2008). Besides making note of their complaints, however, it is difficult to predict within which domains of function a patient may choose to feign illness and disability. For this reason, neuropsychologists are almost always required to examine for disingenuous behavior across multiple cognitive, physical, and emotional domains.

Measures for detecting feigned psychiatric problems are embedded with most standardized psychological inventories (e.g., Lees-Haley Fake Bad Scale; Lees-Haley, English, & Glenn, 1991) and there are a number of standalone tests as well (e.g., Morel Emotional Numbing Test; Morel, 1998). Similarly, evidence of Waddell's Signs (Waddell, McCulloch, Kummel, & Venner, 1980) as well as Bianchini, Greve, and Glynn's (2005) development of a standard for diagnosing Malingered Pain Related Disorder (MPRD) provides the clinician with some basis for judging the authenticity of physical ailments. However, it is difficult to find a single test or assessment protocol that incorporates the complete range of potential domains of feigned responding.

One instrument that is available for use in assessing potential malingering across a variety of psychiatric, cognitive, and physical domains is the Structured Inventory of Malingered Symptomatology (SIMS; Smith & Burger, 1997). The SIMS is composed of 75 self-report, true–false items that are to be answered by individuals 18 and older with at least a fifth-grade reading level. The SIMS is scored by summing the responses on five independent subscales, each containing 15 items, designed to detect the feigning of specific clinical conditions including: psychosis (P), neurologic impairment (NI), amnestic disorders (AM), low intelligence (LI), and affective disorders (AF). In addition, a total score is calculated by summing all of the raw scores (range: 0–75). Although separate cut scores are available for each of the subscales, based on the performance of student simulators, Smith and Burger (1997) initially recommended using a total cut score of 14 in order to identify the presence of malingering. The initial simulation studies used to develop the SIMS have the distinct advantage of having participants who can definitively be identified as malingerers; however, asking simulators to feign symptoms when incentives are minimal or absent often does not generalize to forensic settings where examinees may be attempting to avoid criminal prosecution or receive sizeable financial compensation (Rogers & Cruise, 1998). For that reason, the malingering research field has generally turned to designs using known-groups that compare the responses of suspected or reasonably identified malingerers to those of individuals presumed to be answering honestly. Known-groups designs greatly increase the generalizability of the findings, but are limited by errors in the initial identification of the known malingerers. Nevertheless, by relying on the specific diagnostic criteria associated with definite, probable, or possible malingering (Bianchini et al., 2005; Slick et al., 1999) this potential threat to internal validity can be mitigated to some degree and neuropsychological evaluations that incorporate elements of these established criteria make it somewhat easier to differentiate persons with and without honest intent.

Statistically, the diagnostic utility of response bias measures is established by calculating various classification figures including sensitivity (SE or the probability of a positive test result in persons who are truly malingering) and specificity (SP or the probability of a negative test result in patients who are not actually malingering; Larrabee & Berry, 2007). In general, it is recommended that cut scores favoring specificity over sensitivity be used to avoid false accusations of malingering in examinees with real disabilities (Larrabee, 2009). Although increases in specificity reduce the ability to detect true malingerers, analyses using likelihood ratios have demonstrated that this is easily rectified by administering multiple independent SVTs with high specificity, hopefully across more than one domain of function (Larrabee, 2008). The likelihood of then obtaining a false positive significantly decreases with each subsequent SVT, whereas the ability to detect true positives, assuming that the measures are not strongly correlated, remains relatively unchanged (Larrabee, 2008; Rosenfeld, Sands, & Van Gorp, 2000). The base rate, or frequency of occurrence of malingering, can then be used to accurately calculate Positive Predictive Power (PPP) and Negative Predictive Power (NPP). PPP is the probability of accurately identifying a referral as malingering given a positive test finding, whereas NPP is the probability of accurately classifying a referral as an HR given a negative test finding (Larrabee & Berry, 2007). Unfortunately, the majority of studies attempting to predict malingering calculate PPP and NPP without taking into account the base rate in the population, which often results in erroneous estimates of classification accuracy (Rosenfeld et al., 2000).

To date several different studies have been published in which these classification statistics are used to examine the effectiveness of the SIMS to detect malingering (Table 1). Studies using simulation designs indicate the SIMS has satisfactory diagnostic classification statistics (Edens, Otto, & Dwyer, 1999; Edens, Poythress, & Watkins-Clay, 2007; Jelicic, Hessels, & Merckelbach, 2006; Merckelbach & Smith, 2003; Rogers, Hinds, & Sewell, 1996; Smith & Burger, 1997). As noted earlier, these simulation studies may lack adequate external validity. Several of the known-groups studies have also supported the diagnostic accuracy of the SIMS (Clegg, Fremouw, & Mogge, 2009; Edens et al., 2007; Heinze & Purisch, 2001; Lewis, Simcox, & Berry, 2002; Vitacco, Rogers, Gabel, & Munizza, 2007). However, several of these studies (Lewis et al., 2002;

Table 1. Published studies examining the structured inventory of malingered symptomatology (SIMS)

References	Study design	Sample size		Cut-off	SE	SP	PPP	NPP
		MAL	HR					
Clegg et al. (2009)	Known-groups	35	20	14	1.00	.37	.48	1.00
Edens et al. (1999)	Simulation	196	196	14	.96	.91	.92[a]	.96[a]
Edens et al. (2007)	Known-groups	26	30	14	.85	.40	.55[a]	.75[a]
Edens et al. (2007)	Simulation	29	30	14	.90	.97	.96[a]	.91[a]
Heinze and Purisch (2001)	Known-groups	57	None	13	.87	None	None	None
Jelicic et al. (2006)	Simulation	45	15	16	.91[a]	1.00[a]	1.00[a]	.79[a]
Lewis et al. (2002)	Known-groups	24	31	16	1.00	.61	.54	1.00
Merckelbach and Smith (2003)	Simulation	57	241	16	.93	.98	.90	.98
Rogers et al. (1996)	Simulation	19	34	16	NR	NR	.87	.62
Smith and Burger (1997)	Simulation	NR	NR	14	.96	.88	NR	NR
Vitacco et al. (2007)	Known-groups	21	79	14	1.00	.65	.43	1.00

Notes: MAL = malingerers; HR = honest responders; SE = sensitivity [true positives/(true positives + false negatives)]; SP = specificity [true negatives/(true negatives + false positives)]; PPP = positive predictive power [true positives/(true positives + false positives)]; NPP = negative predictive power [true negatives/(true negatives + false negatives)]; NR = not reported.
[a]Calculated using available information.

Vitacco et al., 2007) were hampered by the difficulty inherent in using the Structured Interview of Reported Symptoms (SIRS; Rogers, Bagby, & Dickens, 1992) as the only measure to diagnose malingering as opposed to the recommended use of multiple sources of data and information as recently outlined by Heilbronner et al. (2009). These studies also selected cut scores that favor sensitivity over specificity to better address their research questions (e.g., screening in potential feigners for further evaluation). As a result, neuropsychologists using these cut scores may erroneously misclassify truly impaired individuals as malingerers. Furthermore, once sensitivity and specificity were established, base rates were different across each of the studies. As a result, the PPP and NPP associated with existing research on the SIMS have a very wide range depending on the study design and participants (.43–1.00 and .62–1.00, respectively). Given these limitations, the current study was designed to broaden the forensic sample used to establish the diagnostic validity of the SIMS while also identifying appropriate cut scores that substantially increase specificity without sacrificing too much sensitivity.

Materials and Methods

Participants

Archival data from January 2005 to June 2009 was reviewed to obtain the psychological test results used in this study. An initial sample of 54 clients, 18 years of age and older with identifiable external incentives were evaluated by a board certified clinical neuropsychologist in a private practice setting. This sample was then reduced by eliminating evaluations in which the SIMS was not administered. In addition, three clients with borderline intelligence were eliminated from the sample because of their higher risk of failing SVTs (Chafetz, 2008) as well as the lack of specific indicators of simulated impairment on measures of mental retardation (for a review, see Almstrom, Wisdom, & Callahan, 2008). Furthermore, a review of the relevant literature suggests traditional measures of cognitive malingering for use with intellectually disabled individuals lack validity (Salekin & Doane, 2009).

The final sample consisted of 33 clients (Table 2) referred for evaluation following a closed head injury (CHI) and/or adverse physical or psychological reactions to trauma (21% and 79%, respectively). The sample comprised 14 men and 19 women ranging in age from 24 to 65 ($M = 44.1$, $SD = 9.7$). Education ranged from 8 to 16 years ($M = 12.4$, $SD = 2.1$) and the length of time between the initial injury and evaluation varied from 18 to 160 months ($M = 48.9$, $SD = 31.4$). Irrespective of alleged referral diagnosis, 60% of the sample consisted of clients who complained of physical, cognitive, and emotional symptoms, whereas another 10% complained of cognitive and either physical or emotional sequela. The remaining 30% complained of physical and/or emotional sequel, but voiced no cognitive concerns. It is important to note that all but one patient voiced complaints in two or more categories of functioning. The majority of the referrals were medicolegal in nature with two exceptions: one referral from the Veterans Administration and one from the State Department of Vocational Rehabilitation. The remainder of the referrals included court ordered independent medical evaluations (31%), defense attorneys (31%), treating physicians (18%), case managers/insurance adjustors (15%), and plaintiff attorneys (5%).

Based on the results of the neuropsychological evaluation, 67% of the sample was diagnosed with either Malingered Neuropsychological Deficit or Malingered Pain Disorder. Of the remaining sample, 9% were diagnosed with poor

Table 2. Means, standard deviations, and ranges for demographic variables

	Definite MAL	Probable MAL	HR	Total
n	5	17	11	33
Age	46.8 ± 10.6 (36–64)	47.4 ± 8.4 (31–65)	37.8 ± 8.9 (24–52)	44.1 ± 9.7 (24–65)
Gender (M/F)	2/3	7/10	5/6	14/19
Education	12.4 ± 3 (8–16)	12.2 ± 1.7 (9–16)	12.8 ± 2.3 (8–16)	12.4 ± 2.1 (8–16)
Referral	4 CHI/1 ψo	2 CHI/15 ψo	1 CHI/10 ψo	7 CHI/26 ψo
MBE	39.6 ± 12.3 (29–59)	55.6 ± 39.8 (18–160)	42.9 ± 20.1 (22–96)	48.9 ± 31.4 (18–160)
SIMS	26 ± 6.5 (20–34)	26.6 ± 8.9 (12–41)	15.3 ± 4.7 (10–23)	22.7 ± 9 (10–41)

Notes: MAL = malingerers; HR = honest responders; CHI = closed head injury; ψo = psychological overlay; MBE = months between injury and evaluation; SIMS = Structured Inventory of Malingered Symptomatology.

psychological adjustment following CHI, 15% were diagnosed with Adjustment and/or Mood Disorder, 6% were diagnosed with Post-traumatic Stress Disorder, and 3% were thought to be suffering from a Somatoform Disorder. The evaluation of one particular client failed to reveal any evidence of physical, cognitive, or emotional problems.

Procedure

Given the range of potential disingenuous behavior, construction of a known-groups research design was achieved by relying on various combinations of response bias measures (Table 3) in conjunction with the criteria of Slick and colleagues (1999) to identify the absence or presence of definite, probable, or possible malingering. Given that a flexible test battery was used for these evaluations, several different response bias measures were selected for use based upon the clients' referral question, their complaint list, and their performance during the evaluation. As a result, not all of these measures were consistently given to each patient. In addition, the response bias measures employed are not psychometrically identical (e.g., varying sensitivity and specificity estimates). Nevertheless, as noted in Table 3, the Test of Memory Malingering (TOMM; Tombaugh, 1996), Dot Counting Test (Boone, Lu, et al., 2002), Word Memory Test (WMT; Green, Iverson, & Allen, 1999), Rey 15-Item Recognition Test (Boone, Salazar, Lu, Warner-Chacon, & Razani, 2002), Rarely Missed Index (RMI; Killgore & DellaPietra, 2000), McGill Pain Questionnaire (Melzack, 1975) Modified Somatic Perception Questionnaire (Main, 1983), Million Visual Analog Scale (Million, Hall, Nilsen, Baker, & Jayson, 1982), Lees-Haley Fake Bad Scale (FBS; Lees-Haley et al., 1991), and the Catastrophizing scale of the Coping Strategies Questionnaire (CSQ; Rosenstiel & Keefe, 1983) were employed at various times. It is beyond the scope of this article to review the extensive literature associated with each of these measures but a thorough review can be found in Block, Gatchel, Deardorff, and Guyer (2003, chap. 5), Boone and Lu (2007), or Grote and Hook (2007).

Following the criteria of Slick and colleagues (1999) and/or Bianchini and colleagues (2005), clients were categorized as: Definite Malingering when their performance on a forced-choice test was significantly worse than would be expected by chance alone; Probable Malingering when their performance was consistent with feigning on any combination of two or more well-validated psychometric tests designed to measure exaggeration/fabrication of deficits AND discrepancies were noted between their reported symptoms and the symptoms observed during the clinical interview or subsequent testing.

All clients evaluated in the current study provided written consent for neuropsychological testing and gave their permission for the collected data to be de-identified and used for research. In addition, all clients were reminded about the forensic boundaries concerning the confidentiality of the assessment and were told they could withdraw from the evaluation at any time. Finally, each patient was informed that effort tests would be embedded in the evaluation and if they were not able to give full effort at any point during the exam, they were to advise the examiner and the testing would be halted with the option of rescheduling at a later date. None of the clients in this study opted for discontinuation of the testing.

Results

Sample Characteristics

Based on the criteria of Slick and colleagues (1999) and/or Bianchini and colleagues (2005), the clients were categorized into three groups: $n = 5$ definite malingers (dMAL), $n = 17$ probable malingerers (pMAL), and $n = 11$ non-malingering or HRs. As a result of the small dMAL sample size, both malingering groups were combined (MAL) for the analyses resulting in 22 highly suspected malingerers (Table 2). No significant differences were found between men and women in the MAL and

Table 3. Means, standard deviations, and range for tests used as criteria for malingering

Test	Subject group		*t*-test
	MAL	HR	
McGill			
n	18	10	
$M \pm SD$ (range)	49.1 ± 8.1 (28–56)	36.3 ± 12.4 (13–53)	3.30**
MSPQ			
n	19	10	
$M \pm SD$ (range)	21.1 ± 4.7 (12–28)	10.4 ± 6.9 (0–25)	4.97**
VA			
n	15	11	
$M \pm SD$ (range)	6.9 ± 2.2 (.79–10)	4.7 ± 2.1 (2–9.5)	2.57*
TOMM			
n	3	0	
$M \pm SD$ (range)	37.3 ± 6.8 (32–45)		
RMI			
n	4	1	
$M \pm SD$ (Range)	134.3 ± 81 (54–204)	204	
Dot			
n	18	7	
$M \pm SD$ (Range)	19.8 ± 12 (7–55)	11 ± 2.7 (8–16)	2.94**
WMT			
n	6	0	
$M \pm SD$ (Range)	24.2 ± 11.1 (4–35)		
Rey			
n	9	1	
$M \pm SD$ (Range)	18.4 ± 6.5 (10–30)	20	
FBS			
n	19	10	
$M \pm SD$ (Range)	30.9 ± 4.1 (22–37)	25.3 ± 6.9 (14–37)	2.36*
CSQ			
n	8	3	
$M \pm SD$ (Range)	30.3 ± 5.3 (19–35)	28.7 ± 18.6 (11–48)	.146

Notes: MAL = malingerers; HR = honest responders; McGill = McGill Pain Questionnaire; MSPQ = Modified Somatic Perception Questionnaire; VA = Million Visual Analog Scale; TOMM = Test of Memory Malingering; RMI = Rarely Missed Index; Dot = Dot Counting Test; WMT = Word Memory Test (lowest scoring trial); Rey-15 = Rey 15-Item Recognition Test; FBS = Lees-Haley Fake Bad Scale; CSQ = Coping Strategies Questionnaire-catastrophizing subtest.
*$p < .05$
**$p < .01$

HR conditions ($\chi^2 = .062$, $p = .80$). Older clients were significantly more likely to be classified as MAL ($M = 47.2$, $SD = 8.7$) when compared with the younger clients ($M = 37.8$, $SD = 8.9$), $F(1, 31) = 8.51$, $p < .05$, but no significant differences were found for years of education between the MAL ($M = 12.2$, $SD = 2$) and the HR group ($M = 12.8$, $SD = 2.3$), $F(1, 31) = .595$, $p = .45$. Likewise, no significant differences were found when looking at the time since injury between the MAL ($M = 52$, $SD = 35.8$) and the HRs ($M = 42.9$, $SD = 20.1$), $F(1, 31) = .601$, $p = .44$.

Several significant differences were found between malingers and HRs when looking at the measures used for categorization (Table 3). More specifically, clients in the MAL group were significantly more likely to exhibit scores consistent with a feigned presentation on the McGill Pain Questionnaire, Modified Somatic Perception Questionnaire, Million Visual Analog Scale, Dot Counting Test, and the Lees-Haley Fake Bad Scale. No significant differences were found between the groups on the Catastrophizing Scale of the Coping Strategies Questionnaire. Independent samples *t*-tests were not conducted on the remaining measures (Test of Memory Malingering, Rarely Missed Index, Word Memory Test, and Rey 15-Item Recognition Test) as a result of inadequate sample sizes.

Discriminant Validity

Table 4 displays the sensitivity and specificity of the SIMS when utilizing specific cut scores. The original (Smith & Burger, 1997) and most widely used cut score of 14 detected 96% of true malingerers; however, this cut score also incorrectly identified 36% of clients as malingering that were actually thought to be responding honestly. Alternatively, a cut score of 24 missed

Table 4. Sensitivity and specificity of the SIMS at detecting malingering using various cut scores

Cut score	Sensitivity	Specificity
≥14	.96	.64
≥21	.68	.73
≥22	.64	.82
≥23	.55	.91
≥24	.55	1.00
≥25	.50	1.00

Table 5. Effect of base rates on the positive predictive power and negative predictive power of the SIMS using a cut score ≥ 24

Base rate	Positive predictive power	Negative predictive power
.10	1	.95
.20	1	.90
.30	1	.84
.40	1	.77
.50	1	.67
.60	1	.60
.70	1	.49

about 45% of true malingerers while accurately classifying 100% of the HRs. Table 5 displays the PPP and NPP of the SIMS at various base rates when using a cut score of 24 that accurately classified 100% of individuals diagnosed as malingering and 77% of clients scoring below 24 as HRs.

Discussion

Because of the breadth of items sampled, the SIMS has the potential to become one of the more valuable response bias measures available; however, no studies in the literature have examined the diagnostic validity of this instrument in a known-groups design involving medicolegal referrals diagnosed as malingering using the gold standard criteria outlined by Slick and colleagues (1999). In addition, the published research studies using the SIMS favored the ability to detect true malingerers while increasing the likelihood of false positives. It is important to note that several of these studies purposefully favored sensitivity to better address their own research questions. Regardless of the rationale, no cut scores designed to maximize specificity in clinical forensic populations have been published in the literature.

The current known-groups, archival study was designed to examine the effectiveness of the SIMS at identifying malingering in clients involved in personal injury lawsuits or disability claims. A battery of SVTs was used to classify the clients into malingering and non-malingering conditions following the established criteria in the literature (Bianchini et al., 2005; Slick et al., 1999). Where sample sizes were large enough to permit statistical analysis, those clients identified as MAL performed significantly worse on all of the measures of malingering when compared with HRs with the exception of the Catastrophizing subtest of the Coping Strategies Questionnaire. When using lower cut scores, the observed sensitivity and specificity of the SIMS in the current study was nearly identical to published findings in the literature (Edens et al., 2007; Lewis et al., 2002; Vitacco et al., 2007). However, in this study, the sensitivity and specificity of the SIMS was calculated at each successive cut score until the rate of false positives was reduced to zero. Using this new cut score, the predictive validity of the SIMS was then calculated using estimates of the malingering base rates found in the literature (Rosenfeld et al., 2000).

The current study found that a cut score of 24 completely eliminates the incidence of false positives. Although this new recommended cut score reduces the ability to detect true malingerers, recent findings indicate that the use of multiple independent SVTs does not increase the rate of false positives (Larrabee, 2008). In addition, it is arguably more important to let a greater number of malingerers in a medicolegal context be successful in their attempts to feign symptoms, so that clients with legitimate concerns can receive the necessary accommodations. With that being said, there are occasions in which cut scores favoring sensitivity are preferred; however, as recommended by Larrabee (2008), near perfect rates of sensitivity can still be achieved through the use of multiple, independent SVTS without increasing the rate of false positives. As a result, the current study recommends using the SIMS in conjunction with other uncorrelated SVTs so that true malingerers are identified without falsely accusing HRs.

There are several limitations of the current study that raises concerns about the generalizability of these findings. Foremost, the small sample sizes observed in each group may have had untoward effects on the diagnostic validity calculations. More specifically, not all possible scores on the SIMS were observed which raises the possibility of imprecise rates of sensitivity and specificity. Another limitation that is important to note is the selection bias concerning when the SIMS was administered. The use of a flexible battery approach in concert with the patient's history and list of presenting complaints resulted in variable measure inclusion. As a result, the SIMS was not always given. This method of selection bias resulted in fewer HRs with available SIMS scores, which in turn artificially raised the base rate of malingering beyond that typically encountered in neuropsychological evaluations (Mittenberg et al., 2002). This study also lacked experimental controls because all of the data used was archival. This unfortunately resulted in discrepant sample sizes on the SVTs used to categorize the clients.

It is recommended that future researchers attempt to replicate the findings of the current study in order to strengthen or refine the suggested cut score of 24 on the SIMS when examining other clinical populations. More specifically, the literature is lacking in studies examining the diagnostic accuracy of the SIMS in populations diagnosed with a somatoform disorder. In fact, it may be inherently difficult to differentiate between somatoform disorders and malingering because both involve the endorsement of a broad range of symptoms in which there is no identifiable physical cause. As a result, it is quite possible that clients presenting with a legitimate somatoform disorder will appear to be malingering on the SIMS as well as other SVTs.

Additional studies should also be completed with other normative groups such as learning disabilities or those with MCI in an effort to broaden (or limit) the applicability of the SIMS in these populations. Additional analyses designed to refine and better identify appropriate cut scores for each of the individual subtests of the SIMS would also be useful. However, until or unless a typical pattern of "malingering" emerges in the literature, sampling self-report measures across cognitive, physical, and emotional domains in a single test appears to have benefit for improving the accuracy of forensic assessments.

Conflict of interest

None declared.

References

Almstrom, C. M., Wisdom, N. M., & Callahan, J. L. (2008). Beyond full scale IQ: A new WAIS-III indicator of mental retardation. *Journal of Scientific Psychology*, *3*, 9–14.

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

Bianchini, K. J., Greve, K. W., & Glynn, G. (2005). On the diagnosis of malingered pain-related disability: Lessons from cognitive malingering research. *The Spine Journal*, *5*, 404–417.

Binder, L. M., & Rohling, M. L. (1996). Money matters: A meta-analytic review of the effects of financial incentives on recovery after closed-head injury. *American Journal of Psychiatry*, *153*, 7–10.

Block, A. R., Gatchel, R. J., Deardorff, W. W., & Guyer, R. D. (2003). *The Psychology of Spine Surgery*. Washington, DC: American Psychological Association.

Boone, K. B. (Ed.) (2007). *Assessment of feigned cognitive impairment: A neuropsychological perspective*. New York, NY: Guilford.

Boone, K. B., & Lu, P. H. (2007). Non-forced-choice effort measures. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (pp. 27–43). New York, NY: Oxford University Press.

Boone, K. B., Lu, P., Back, C., King, C., Lee, A., Philpott, L., et al. (2002). Sensitivity and specificity of the Rey Dot Counting Test in patients with suspect effort and various clinical samples. *Archives of Clinical Neuropsychology*, *17*, 625–642.

Boone, K. B., Salazar, X., Lu, P., Warner-Chacon, K., & Razani, J. (2002). The Rey 15-Item Recognition Trial: A technique to enhance sensitivity of the Rey 15-Item Memorization Test. *Journal of Clinical and Experimental Neuropsychology*, *24*, 561–573.

Bush, S. S., Ruff, R. M., Troster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., et al. (2005). Symptom validity assessment: Practice issues and medical necessity. *Archives of Clinical Neuropsychology*, *20*, 419–426.

Chafetz, M. D. (2008). Malingering on the social security disability consultative exam: Predictors and base rates. *The Clinical Neuropsychologist*, *22*, 529–546.

Clegg, C., Fremouw, W., & Mogge, N. (2009). Utility of the Structured Inventory of Malingered Symptomatology (SIMS) and the Assessment of Depression Inventory (ADI) in screening for malingering among outpatients seeking to claim disability. *The Journal of Forensic Psychiatry & Psychology*, *20*, 239–254.

Edens, J. F., Otto, R. K., & Dwyer, T. (1999). Utility of the Structured Inventory of Malingered Symptomatology in identifying persons motivated to malinger psychopathology. *Journal of the American Academy of Psychiatry and the Law*, *27*, 387–396.

Edens, J. F., Poythress, N. G., & Watkins-Clay, M. M. (2007). Detection of malingering in psychiatric unit and general population prison inmates: A comparison of the PAI, SIMS, and SIRS. *Journal of Personality Assessment*, *88*, 33–42.

Green, P., Iverson, G. L., & Allen, L. (1999). Detecting malingering in head-injury litigation with the Word Memory Test. *Brain Injury*, *13*, 813–819.

Grote, C. L., & Hook, J. N. (2007). Forced-choice recognition tests of malingering. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (pp. 44–79). New York, NY: Oxford University Press.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., & Millis, S. R. Conference Participants (2009). American Academy of Clinical Neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, *23*, 1093–1129.

Heinze, M. C., & Purisch, A. D. (2001). Beneath the mask: Use of psychological tests to detect and subtype malingering in criminal defendants. *Journal of Forensic Psychology Practice*, *4*, 23–52.

Holsinger, T., Steffens, D. C., Phillips, C., Helms, M. J., Havlik, R. J., Breitner, J. C. S., et al. (2002). Head injury in early adulthood and the lifetime risk of depression. *Archives of General Psychiatry*, *59*, 17–22.

Iverson, G. L., & Binder, L. M. (2000). Detecting exaggeration and malingering in neuropsychological assessment. *Journal of Head Trauma Rehabilitation*, *15*, 829–858.

Jelicic, M., Hessels, A., & Merckelbach, H. (2006). Detection of feigned psychosis with the Structured Inventory of Malingered Symptomatology (SIMS): A study of coached and uncoached simulators. *Journal of Psychopathology and Behavior Assessment*, *28*, 19–22.

Killgore, W. D. S., & DellaPietra, L. (2000). Using the WMS-III to detect malingering: Empirical validation of the Rarely Missed Index (RMI). *Journal of Clinical and Experimental Neuropsychology*, *22*, 761–771.

Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist*, *17*, 410–425.

Larrabee, G. J. (2008). Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios. *The Clinical Neuropsychologist*, *22*, 666–679.

Larrabee, G. J. (2009, February). *Assessment of malingered neuropsychological deficits. Workshop presented at the annual meeting of the International Neuropsychological Society*, Atlanta, GA.

Larrabee, G. J., & Berry, D. T. R. (2007). Diagnostic classification statistics and diagnostic validity of malingering assessment. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (pp. 14–26). New York, NY: Oxford University Press.

Lees-Haley, P. R., English, L. T., & Glenn, W. J. (1991). A Fake Bad Scale on the MMPI-2 for personal injury claimants. *Psychological Reports*, *68*, 203–210.

Lewis, J. L., Simcox, A. M., & Berry, D. T. R. (2002). Screening for feigned psychiatric symptoms in a forensic sample by using the MMPI-2 and the Structured Inventory of Malingered Symptomatology. *Psychological Assessment*, *14*, 170–176.

Main, C. J. (1983). The Modified Somatic Perception Questionnaire (MSPQ). *Journal of Psychosomatic Research*, *27*, 503–514.

Melzack, R. (1975). The McGill Pain Questionnaire: major properties and scoring methods. *Pain*, *1*, 277–299.

Merckelbach, H., & Smith, G. P. (2003). Diagnostic accuracy of the Structured Inventory of Malingered Symptomatology (SIMS) in detecting instructed malingering. *Archives of Clinical Neuropsychology*, *18*, 145–152.

Million, R., Hall, W., Nilsen, K. H., Baker, R. D., & Jayson, M. I. (1982). Assessment of the progress of the back-pain patient. *Spine*, *7*, 204–212.

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*, 1094–1102.

Morel, K. R. (1998). Development and preliminary validation of a forced-choice test of response bias for post-traumatic stress disorder. *Journal of Personality Assessment*, *70*, 299–314.

Nampiaparampil, D. E. (2008). Prevalence of chronic pain after traumatic brain injury: A systematic review. *The Journal of the American Medical Association*, *300*, 711–719.

Rogers, R. R. (1997). Introduction. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (2nd ed., pp. 1–19). New York, NY: Guilford.

Rogers, R. R., & Cruise, K. R. (1998). Assessment of malingering with simulation designs: Threats to external validity. *Law and Human Behavior*, *22*, 273–285.

Rogers, R. R., Hinds, J. D., & Sewell, K. W. (1996). Feigning psychopathology among adolescent of offenders: Validation of the SIRS, MMPI-A, and SIMS. *Journal of Personality Assessment*, *67*, 244–257.

Rosenfeld, B., Sands, S. A., & Van Gorp, W. G. (2000). Have we forgotten the base rate problem? Methodological issues in the detection of distortion. *Archives of Clinical Neuropsychology*, *15*, 349–359.

Rosenstiel, A. K., & Keefe, F. J. (1983). The use of coping strategies in chronic low back pain patients: Relationship to patient characteristics and current adjustment. *Pain*, *17*, 33–44.

Salekin, K. L., & Doane, B. M. (2009). Malingering intellectual disability: The value of available measures and methods. *Applied Neuropsychology*, *16*, 105–113.

Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, *13*, 545–561.

Smith, G. P., & Burger, G. K. (1997). Detection of malingering: Validation of the Structured Inventory of Malingered Symptomatology (SIMS). *Journal of the American Academy of Psychiatry and the Law*, *25*, 183–189.

Sweet, J. J., King, J. H., Malina, A. C., Bergman, M. A., & Simmons, A. (2002). Documenting the prominence of forensic neuropsychology at national meetings and in relevant professional journals from 1990 to 2000. *The Clinical Neuropsychologist*, *16*, 481–494.

Tombaugh, T. N. (1996). *Test of Memory Malingering (TOMM)*. New York, NY: Multi-Health Systems.

Vitacco, M. J., Rogers, R., Gabel, J., & Munizza, J. (2007). An evaluation of malingering screens with competency to stand trial patients: A known-groups comparison. *Law and Human Behavior*, *31*, 249–260.

Waddell, G., McCulloch, J. A., Kummel, E., & Venner, R. M. (1980). Nonorganic physical signs in low-back pain. *Spine*, *5*, 117–125.