



Structured Inventory of Malingered Symptomatology: a Psychometric Review

Robert D. Shura^{1,2,3} · Anna S. Ord^{1,2,3,5} · Mark D. Worthen⁴

Received: 30 August 2021 / Accepted: 31 October 2021 / Published online: 25 November 2021
This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2021

Abstract

The Structured Inventory of Malingered Symptomatology (SIMS) is a stand-alone symptom validity test. The 75-item, true/false measure samples feigned symptoms across five subscales: Psychosis, Neurologic Impairment, Amnestic Disorders, Low Intelligence, and Affective Disorders. Initial psychometrics reported in the manual were adequate, and since publication, dozens of studies have been published on the SIMS, with practice surveys indicating the measure as one of the most commonly used SVTs. Several SIMS short forms, alternate scales, and translations have been published. The SIMS purports to measure feigned self-reported symptoms. Several studies demonstrate convergent validity for the SIMS, as well as incremental validity when compared to clinical judgment based on interview and record review alone. Cutoff scores were quite variable across studies, but when using common cutoff scores, the SIMS does not reliably distinguish feigned psychopathology from severe manifestations of genuine psychiatric illness. This lack of robust discriminant validity means that evaluators need to consider higher SIMS cutoff scores in certain circumstances, which we describe herein. Van Impelen et al. published a meta-analysis on the SIMS in 2014; an updated diagnostic accuracy table is provided here, including new research since the prior review. We conclude by highlighting strengths, weaknesses, and areas of future exploration with the SIMS.

Keyword Structured Inventory of Malingered Symptomatology; SIMS; Symptom validity test; Exaggerated/feigned symptom

Background and Conditions of Use

The Structured Inventory of Malingered Symptomatology (SIMS; Widows & Smith, 2005) is a 75-item, freestanding symptom validity test (SVT). Despite the test name including the term malingering, the SIMS manual states that the purpose of the measure is to “assess symptoms of both feigned psychopathology and cognitive function”

(Widows & Smith, 2005, p. 4), not malingering per se. *Malingering* as commonly used and defined in the *DSM-5* involves “intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by secondary incentives” (American Psychiatric Association, 2013, p. 726): as the SIMS does not assess intentionality or external incentives, it is best described as a measure of negative response bias or symptom feigning (exaggeration or fabrication of symptoms). The SIMS is administered on a carbonless booklet by marking directly on the form, via local administration software, or through the online *PARi-Connect* platform. Items are true–false in format, and the measure takes about 15 min to complete. It was designed for use in adults (18 years of age and older) with at least 5th grade reading level (Widows & Smith, 2005). The total score includes all items and tends to be most researched, but the SIMS also contains five subscales of 15 items each: Psychosis (P), Neurologic Impairment (NI), Amnestic Disorders (AM), Low Intelligence (LI), and Affective Disorders (AF). Practice surveys have identified the SIMS as one of the most commonly used SVTs (Martin et al., 2015).

✉ Robert D. Shura
Robert.Shura2@va.gov

¹ VA Mid-Atlantic Mental Illness Research, Education, and Clinical Center (MA-MIRECC), Salisbury, NC, USA

² Research & Academic Affairs Service Line, Salisbury Veterans Affairs Medical Center, Salisbury, NC, USA

³ Department of Neurology, Wake Forest School of Medicine, Winston-Salem, NC, USA

⁴ Health West, Inc. (Federally Qualified Health Center), Pocatello, ID, USA

⁵ Regent University, VA Beach, USA

SIMS items were developed in several phases. During the first phase, 200 items were generated based on the review of literature that described qualitative characteristics of malingerers (Widows & Smith, 2005). During the next stage, nine experienced clinical psychologists were asked to rate the initial pool of 200 items and to classify each item into different conditions: low intelligence, affective disorder, neurologic impairment, psychosis, amnesic disorder, another category, or some combination of these categories (Smith & Burger, 1997; Widows & Smith, 2005). Items that received at least a two-thirds (67%) agreement rate were placed into one of five aforementioned conditions in rank order of their percentage agreement rate (Smith & Burger, 1997; Widows & Smith, 2005). The items that did not meet the 67% agreement rate were discarded. The final set contained 75 items (15 items per each of five scales) with interrater reliability values ranging from 0.76 (Neurologic Impairment) to 0.95 (Affective Disorders) with a mean interrater reliability coefficient of 0.84. The initial validation sample included 476 undergraduate students, with a mean age of 24.43 years and who were 71% female and 89.7% White. A simulation design was used to establish initial scores. The SIMS scales have been shown to demonstrate acceptable internal consistency with reliability coefficients ranging from 0.80 to 0.88 (Smith & Burger, 1997).

Initial principal component analysis reported in the SIMS manual (Widows & Smith, 2005) indicated a four-factor solution, with most of the NI items loading on the P subscale; however, NI was kept as a separate subscale given the good Cohen's kappa score. In addition to the total score and five original subscales, Rogers et al. (2014) developed two new scales using a sample of psychiatric inpatients. In contrast to the original subscales that screen for feigning by symptom domain (e.g., memory versus psychotic symptoms), Rogers et al. developed their new scales based on detection methods. Fifteen items that were rarely endorsed by genuine responders but were commonly endorsed by simulated feigners comprise the Rare Symptoms (RS) scale. Thirteen item pairs that were highly intercorrelated in the feigning group and uncorrelated in the genuine group constitute the Symptom Combinations (SC) scale. The items on the SC scale do not necessarily reflect pairs of similar symptoms as sampled by two separate items; rather, the empirical method for deriving resulted in a variety of symptoms across all five of the original SIMS subscales. Reliabilities for the two new scales were excellent (RS) and moderate (SC). Rogers et al. (2014) recommended a cutoff score of > 6 for both new scales.

Short Forms

Completion of the SIMS is less burdensome than long multi-scale self-report measures with embedded SVTs; however,

several studies using various methodologies identified potentially viable short forms of the SIMS. The earliest (Malcore et al., 2015) utilized a two-step process for item reduction: first, all items with a corrected correlation to the total score of less than $r=0.30$ were removed ($n=37$); then, items were removed if 95% of respondents endorsed the same answer, indicating lack of variability. Finally, the remaining two items of the LI subscale and one more item from the P subscale were dropped, leaving a 36-item short form (SIMS-SF) including four of the five subscales (of note, the short form appears to include 35 items from the description and others' reference to the scale). Initial area under the curve (AUC) analyses suggested that the abbreviated total score, and even the abbreviated NI and AM scores, was comparable to the original total score when compared to FBS and RBS from the Minnesota Multiphasic Personality Inventory 2nd edition (MMPI-2). A separate study utilized machine learning to identify a SIMS short form in a sample of 132 subjects with adjustment disorder who completed a psychological evaluation in the context of a lawsuit (Orrù et al., 2021). Subjects were divided into the following groups based on three criteria: consistent, accentuators (i.e., symptom exaggerators), and producers (i.e., symptom fabricators). The resulting short forms included an 8-item version and a 10-item version, which were compared to both the Malcore et al. (2015) SIMS-SF and the full SIMS, with all models showing potential promise psychometrically; however, the items contained in those models were not listed in the manuscript.

More recently, Spencer et al. (2021) created the SIMS for Neuropsychological Settings (SIMS-NS) scale comprising 11 NI items and 8 AM items, retained based on correlations to MMPI-2 Restructured Form (MMPI-2-RF) SVTs agreed to be most relevant: AM items were compared to RBS and NI items were compared to FBS-r and Fs. The authors then compared the original SIMS, the new SIMS-NS, Rogers' RS and SC scales, and the Malcore SIMS-SF on their ability to identify those who produce invalid scores on performance validity tests (PVTs). AUCs were comparable for SIMS Total (0.76), SIMS-NS (0.74), SIMS-SC (0.72), and SIMS-SF (0.76). Although these studies are promising for various short forms, different subscales, and using SIMS scores to screen for possible invalid performance, additional research in this area is warranted.

Appropriate Age Range and Language

The SIMS manual states that the measure is intended for adults aged 18 and older (Widows & Smith, 2005); however, a simulation study by Rogers et al. (1996) evaluated the SIMS in a sample of 53 adolescents ages 14–17. Although the total score had a PPP of 0.87 at > 14 (NPP = 0.62), the AF subscale showed improved accuracy at > 5 (PPP = 0.91, NPP = 0.70). For that sample, the total score was raised to

a cutoff of > 40 to optimize accuracy, leading to $PPP = 0.49$ and $NPP = 0.94$. No studies were found on children, though at the content level, some items would not apply to children (e.g., driving items, some knowledge items on LI) and the 5th grade reading level would also create lower bounds regarding age. Other than age, the SIMS has been validated in a variety of contexts (see Table 1) including research, clinical, and forensic.

The SIMS has also been studied in several countries other than the USA, such as Spain, Germany, Italy, Switzerland, and Netherlands, as well as several other languages, including Spanish, German, Dutch, and Turkish (Ardic et al., 2019; Giromini et al., 2018; González Ordi et al., 2012; Nijdam-Jones & Rosenfeld, 2017). Related to language and using a modified Dutch version of the SIMS, a small study with asylum seekers (van der Heide & Merckelbach, 2016) found that although those with poor language proficiency endorsed more SIMS items than those with intermediate or good proficiency, this effect was not statistically significant in two different sub-samples. In contrast, SIMS scores were more related to incentives to malingering than to language proficiency. In another study incorporating cultural factors into assessing the SIMS, Boskovic et al. (2017) found that mental health professionals from 22 countries rated items on the SIMS Short Form as less plausible compared to a measure of common psychiatric symptoms (Brief Symptom Inventory-18), regardless of whether the rater was from a Western or non-Western country; however, SIMS item plausibility was rated as not significantly different from a measure of dissociative symptoms, suggesting caution when using the SIMS with those reporting dissociative symptoms.

Meta-Analysis and Practice Surveys

van Impelen et al. (2014) completed a meta-analysis of the SIMS, specifically at cutoff scores of > 14 and > 16 . Although sensitivity was adequate, specificity was less satisfactory and highly variable across various samples. Diagnostic accuracy of the SIMS subscales was also examined and generally adequate, though the P and LI scale were substantially less accurate than the other subscales. The authors recommend a variety of cutoff scores based on different situations, with > 16 recommended for screening purposes and a cutoff of > 24 for clinical populations or more certain conclusions. In addition to meta-analysis, the SIMS has been discussed in published practice surveys. For example, Dandachi-FitzGerald and Merckelbach (2013) completed a survey of neuropsychologists ($N = 515$) in six European countries and reported that SIMS was used by 13.2% of their sample, second to only the MMPI-2. Similarly, Martin et al. (2015) conducted a survey of North American neuropsychologists ($N = 316$) and reported that 10.1% of their sample used the SIMS, which was the most commonly used

stand-alone symptom validity measure; only the MMPI-2/RF and Personality Assessment Inventory (PAI; Morey, 2007) were more commonly used.

Convergent and Incremental Validity

As Kelley (1927, p. 14) explained, “the problem of validity is that of whether a test really measures what it purports to measure.” As noted, the SIMS purports to measure malingered symptomatology, where *symptomatology* means psychopathological symptoms and *malingered* means—contrary to the standard definition—inaccurate symptom endorsement due to significant exaggeration or fabrication, whether or not such behavior arises from a conscious intention to achieve a specific outcome. Thus, the SIMS does not actually assess malingering as the word is commonly defined,¹ a point made convincingly by van Impelen et al. (2014, p. 1337). The construct the SIMS purports to measure is feigned psychological and cognitive symptoms.²³ Despite including scales for feigned memory complaints and low intellectual function, the SIMS does not appear sensitive to underperformance on cognitive tests as measured by performance validity tests, based on the few studies available (Alwes et al., 2008; Dandachi-FitzGerald et al., 2011).

Does the SIMS actually measure feigned symptomatology? The short answer is a qualified yes. As a screening instrument, the SIMS efficiently and effectively *rules out* feigning, i.e., identifying evaluatees who do not require further symptom validity assessment.⁴ Also, as a screening test, the SIMS effectively identifies individuals requiring comprehensive feigning assessment, provided evaluators understand its limitations. Specifically, while several studies have demonstrated convergent validity for the SIMS (e.g., Clegg et al., 2009; Edens et al., 1999; Heinze & Purisch, 2001; Lewis et al., 2002; Merten et al., 2020), some of these same studies

¹ By “commonly defined” we mean “malingering” as a quasi-diagnostic label (e.g., World Health Organization, 2021), and the standard dictionary definition, e.g., “to feign or produce physical or psychological symptoms to obtain financial compensation or other reward” (Oxford University Press, 2018).

² Feigning is defined as “deliberate fabrication or gross exaggeration of psychological or physical symptoms, without any assumptions about its goals” (Rogers, 2018, p. 6). Of note, in some cases, we use terminology by original authors given the increased precision, such as participants specifically instructed to simulate malingering.

³ In some instances, an assessment method might be able to differentiate *exaggerated* from *fabricated* symptoms, but the SIMS has not exhibited this capability to date.

⁴ Provided the evaluator selects an appropriate cutoff score, e.g., ≤ 16 for populations where exaggeration/feigning prevalence is high (Lewis et al. 2002) or ≤ 23 if the exaggeration/feigning base rate is relatively low.

Table 1 Summary of classification accuracy studies for various cutoffs of the SIMS total score

Citation	SIMS cutoff score	SN	SP	Classification accuracy	BR/PPP/NPP	Sample size	Population	Context	Design	External criterion
Alwes et al. (2008)	> 16	0.75–0.96	0.60–0.67	NR	BR = 0.30 PPP = 0.44–0.55 NPP = 0.85–0.97	308	Claimants	Forensic	Known groups	LMT, SIRS, TOMM, VSVT
Benge et al. (2012)	> 14	0.76	0.55	NR	BR = 0.25 PPV = 0.36 NPV = 0.87	120	Veteran inpatients	Clinical	Known groups	Clinical diagnosis
	> 16	0.71	0.69	NR	BR = 0.25 PPV = 0.43 NPV = 0.88					
	> 19	0.59	0.83	NR	BR = 0.25 PPV = 0.53 NPV = 0.86					
	> 23	0.39	0.90	NR	BR = 0.25 PPV = 0.55 NPV = 0.81					
Clegg et al. (2009)	> 14	1.00	0.37	HR = 0.76	BR = NR PPP = 0.48 NPP = 1.00	116	SSDI claimants/ community volunteers	Mixed: Forensic/ Research	Mixed: Known groups/Simula- tion	SIRS
Dandachi-FitzGerald and Merkelbach (2013)	> 16	0.91	1.00	NR	NR	62	Students	Research	Simulation	N/A
Edens et al. (1999)	> 14	0.96	0.91	NR	BR = NR PPP = 0.92 ¹ NPP = 0.96 ¹	196	Students	Research	Simulation	N/A
Edens et al. (2007)	> 14	0.85 ²	0.40 ²	OCC = 0.77	Known groups: PPP = 0.55 ¹ NPP = 0.75 ¹	115	Prison inmates	Mixed: Forensic/ Research	Mixed: Known groups/Simula- tion	Clinical judgment*
	> 14	0.90 ³	0.97 ³		Simulation: PPP = 0.96 ¹ NPP = 0.91 ¹					
Giger et al. (2010)	> 14	0.80 ⁴	0.95	OCC = 0.85	NR	60	Mixed: general population/ military	Research	Simulation	N/A
Grant et al. (2020)	> 14	0.52	1.00	0.57	NR	75	Students	Research	Simulation	N/A
	> 16	0.36	1.00	0.68	NR					
Graue et al. (2007)	> 16	0.88	0.23	HR = 0.55	NR	61	Individuals with mild MR ⁵ and community volunteers	Research	Simulation	N/A

Table 1 (continued)

Citation	SIMS cutoff score	SN	SP	Classification accuracy	BR/PPP/NPP	Sample size	Population	Context	Design	External criterion
Harris and Metz (2021)	> 14	NR	0.75	NR	NR	110	Neuropsychology outpatients	Clinical	Known groups	Clinical judgment
Heinz and Purisch (2001)	> 13	0.87	NR	HR=0.87	NR	57	Psychiatric inpatients	Forensic	Known groups	Clinical judgment
Jelicic et al. (2006)	> 16	0.91 ⁴	1.00 ⁴	0.93 ⁴	BR=NR PPP=1.00 ¹ NPP=0.79 ¹	60	Students	Research	Simulation	N/A
Jelicic et al. (2007a, 2007b)	> 14	0.90	1.00	0.93 ⁴	NR	90	Students	Research	Simulation	N/A
Jelicic et al. (2007a, 2007b)	> 16	0.98 ⁴	1.00	0.99 ⁴	NR	61	Students	Research	Simulation	N/A
Jelicic et al. (2011)	> 16	0.89 ⁴	1.00	0.93 ⁴	NR	86	Mixed: general population/students	Research	Simulation	N/A
Jelicic et al. (2013)	> 16	0.92	1.00	0.96 ⁴	NR	47	Mixed: MH professionals/students	Research	Simulation	N/A
Lewis et al. (2002)	> 16	1.00	0.61	HR=0.73	BR=0.31 PPP=0.54 NPP=1.00	55	MH outpatients	Forensic	Known Groups	SIRS
Merkelbach and Collaris (2012)	> 16	NR	NR	AUC=0.96	NR	90	Students	Research	Simulation	N/A
Merkelbach and Smith (2003)	> 16	0.93	0.98	NR	BR=NR PPP=0.90 NPP=0.98	298	Students	Research	Simulation	N/A
Parks et al. (2017)	> 14	0.81	NR	NR	NR	78	Students	Research	Simulation	N/A
	> 16	0.71								
	> 19	0.58								
	> 24	0.42								
Puente-López et al. (2020)	> 14	0.55 – 0.60	0.79 – 0.98	AUC=0.73–0.94	BR=0.30 PPP=0.53–0.91 NPP=0.81–0.85	650	Mixed: general population/MVA patients	Mixed: Clinical/Research	Mixed: simulation and known groups	Clinical judgment
	> 16	0.52 – 0.55	0.83 – 0.98	AUC=0.73–0.94	BR=0.30 PPP=0.59–0.94 NPP=0.80–0.84					
Rogers et al. (1996)	> 16	0.70 ⁶	0.83 ⁶	NR	BR=NR PPP=0.87 NPP=0.62	53	Adolescents in court-ordered MH treatment	Research	Simulation	N/A

Table 1 (continued)

Citation	SIMS cutoff score	SN	SP	Classification accuracy	BR/PPP/NPP	Sample size	Population	Context	Design	External criterion
Rogers et al. (2005)	> 16	NR	NR	NR	NR	65	Students	Research	Simulation	N/A
Rogers et al. (2014)	> 14	0.93	0.28	OCC=0.60	BR=0.25 PPP=0.30 NPP=0.92	107	MH inpatients	Research	Simulation	SIRS
Smith and Burger (1997)	> 44	0.60	0.98	OCC=0.79	BR=0.25 PPP=0.92 NPP=0.88	476	Students	Research	Simulation	N/A
Vitacco et al. (2007)	> 14	1.00	0.65	HR=0.72	BR=0.21 PPP=0.43 NPP=1.00	100	MH outpatients	Forensic	Known groups	SIRS
Wisdom et al. (2010)	> 13 > 22	0.96 0.55	0.64 0.91	NR NR	NR NR	33	TBI outpatients	Forensic	Known groups	Clinical judgment based on criteria used by Slick et al (1999) and Bianchini et al. (2005)
	> 23	0.55	1.00	NR	BR=0.30 PPP=1.00 NPP=0.84					

Superscript: 1 = These numbers were retrieved from Wisdom et al. (2010), as this information was not available in the original article, but was calculated by Wisdom et al. (2010) based on available information. 2 = These values were reported for the comparison of known groups (SN for “suspected malingers” and SP for “patients”; Edens et al., 2007). 3 = These values were reported for the comparison of simulation groups (SN for “simulators” and SP for “controls”; Edens et al., 2007). 4 = These values were calculated based on information available in each article. 5 = MR = mental retardation; this diagnosis was specified in the article by Graue et al. (2007). This terminology has been retained in this article because although this diagnosis is no longer included in the DSM-5, this terminology was used at the time of the article publication in 2007, and this disorder had specific diagnostic criteria that are different from the currently accepted diagnosis of Intellectual Disability. 6 = These values were retrieved from van Impelen et al. (2014), as this information was not available in the original article, but was calculated by van Impelen et al. (2014) based on available information

SIMS Structured Inventory of Malingered Symptomatology. SV sensitivity, SP specificity, PPP positive predictive power, BR presumed or estimated base rate (prevalence) of exaggeration/feigning, BR base rate used to calculate PPP and NPP (when multiple BR values were presented in the study, the authors selected BR that was closest to 0.3 as this rate is in the range of global expected SVT failure rate across various populations [Dandachi-FitzGerald et al., 2020; Larrabee, 2003; Martin & Schroeder, 2020]), NR not reported, LMT Letter Memory Test, SIRS Structured Interview of Reported Symptoms, TOMM Test of Memory Malingering, VSVT Victoria Symptom Validity Test, N/A not applicable, MH mental health, HR hit rate, AUC area under the curve, OCC overall correct classification, MH mental health, TBI traumatic brain injury forensic = medico-legal contexts (disability evaluation, etc.), clinical context is part of routine health care, research for research only (not clinical or forensic)

have shown that to some extent the SIMS may measure genuine psychopathology (e.g., Ord et al., 2021); thus, evidence for the test's discriminant validity is only moderate. Along these lines, the SIMS may misidentify patients exhibiting marked apathy (Dandachi-FitzGerald et al., 2020), patients with alexithymia (Merckelbach, et al., 2018), patients with schizophrenia (Peters et al., 2013), veterans with PTSD (Wolf et al., 2020), and inpatients with extensive trauma histories (Rogers, et al., 2014), although conversely, the SIMS did not misidentify patients with Korsakoff Amnesia in one study (Oudman et al., 2020).

If evaluators understand that discriminant validity for the SIMS is limited, they can minimize false positive findings by determining a cutoff score, in advance, that takes into account the estimated feigning base rate and the severity of genuine psychopathology typically seen at the evaluation site. For example, imagine a psychologist conducting clinical assessments at a long-term residential treatment program for patients.

- diagnosed with severe substance use disorders;
- who endured multiple adverse childhood events, most of whom are diagnosed with PTSD;
- and who by and large are neither involved in personal injury litigation nor seeking any sort of disability benefit.

The psychologist at such a facility would want to assume a relatively low feigning prevalence rate and would want to take into account research showing that individuals with PTSD, for example, produce substantially higher validity scale scores on the MMPI-2 (Franklin et al., 2002; Frueh et al., 2000) and MMPI-2-RF (Goodwin et al., 2013) compared to other clinical groups. Consequently, the psychologist would probably want to use a SIMS cutoff score of ≤ 23 to *rule out* the need for additional symptom validity assessment. Thus, determining in advance, an evidence-based cutoff score will improve the odds that the SIMS measures feigned psychopathology rather than measuring genuine psychiatric illness.

Convergent Validity

To demonstrate an instrument's convergent validity, researchers identify established operationalizations (measures) of the desired construct, and then determine if the instrument in question correlates highly with those established measures. For symptom validity tests (SVTs), investigators usually compare an SVT with other SVTs already shown to exhibit good reliability and validity. The SIMS manual (Widows & Smith, 2005) refers to several studies using an analog (simulated feigning) research design with undergraduates to assess validity of the SIMS scores. Analyses comparing honest responders and simulator feigning groups have revealed that individuals in simulating groups

produced consistently higher SIMS total scores than those in the honest responding groups ($p < 0.01$). The manual also references several early studies on convergent and divergent validity (included below), but specifically includes results from a small study ($n = 20$, cited as via personal communication) using a disability sample and found the SIMS total score highly correlated with other measures of response bias: PAI NIM $r = 0.94$, M-FAST Total $r = 0.93$, TOMM $r = -0.89$ to -0.91 (Widows & Smith, 2005). Of note, despite the high correlations with the TOMM, another study from the Netherlands (Dandachi-FitzGerald et al., 2011) using a clinical sample ($N = 183$) demonstrated a far lower correlation ($r = -0.22$) between the Dutch version of the SIMS and the Amsterdam Short Term Memory test, a different performance validity measure. Further, unpublished analyses by Smith and Burger (1997) described in the SIMS manual (Widows & Smith, 2005) demonstrated that SIMS total score was shown to have high correlations with the *F* scale ($r = 0.84$) and *F-K* index ($r = 0.81$) on the MMPI. Correlations of SIMS subscales with MMPI *F* scale were moderate to strong ranging from 0.53 (AM and LI) to 0.84 (P). Similarly, correlations of SIMS subscales with the *F-K* index of the MMPI were moderate (0.47 for LI and 0.49 for AM) to strong (0.83 for P). Additionally, SIMS total score was significantly correlated with the Insanity subscale ($r = -0.71$) of the malingering scale and with the 16PF Faking Bad Scale ($r = 0.45$).

Heinze and Purisch (2001) conducted a study involving 57 men suspected of feigning incompetence to stand trial and reported that SIMS total score was moderately or strongly correlated with other commonly used measures of symptom validity: the validity scales of the MMPI-2 (correlations ranged from -0.47 to 0.50) and scales of the Structured Interview of Reported Symptoms (SIRS; Rogers et al., 1992) with correlations ranging from 0.43 to 0.80. Similarly, using the MMPI-2, Lewis et al. (2002) in their study of 64 men undergoing pretrial forensic assessments found that the SIMS total score had the second-best effects size when comparing honest responders to feigners ($d = 3.0$), second only to MMPI-2 *Fb* ($d = 3.6$).

Edens et al. (2007) compared the ability of the SIMS, the SIRS, and the PAI to detect malingering among prison inmates ($N = 115$) in four subsamples of inmates: (a) a group instructed to malingering, (b) a group of suspected malingerers who were identified by psychiatric staff, (c) a control group selected from general population of inmates, and (d) a group of psychiatric patients. They reported that intercorrelations among measures for the total sample were quite high. Specifically, SIMS total scores were correlated strongly with the SIRS scores ($r = 0.81$), the NIM subscale of the PAI ($r = 0.84$), and moderately with two other PAI indices: MAL $r = 0.68$ and RDF $r = 0.45$. The SIMS total score correlated to the SIRS total score at a far lesser degree ($r = 0.54$) in a

residential treatment sample ($n=41$) of veterans diagnosed with PTSD (Freeman et al., 2008). Thus, in general, concurrent validity is supported when the SIMS is compared to a variety of other SVTs.

Incremental Validity and Incremental Utility

If an evaluator routinely administers one of the MMPI instruments (MMPI-2, MMPI-2-RF, or MMPI-3) or the PAI to screen for exaggeration/feigning, they might wonder if the administration of the SIMS would be useful. In other words, would the SIMS improve the ability to identify feigning? Unfortunately, few studies have examined incremental validity of the SIMS, and definitive evidence supporting incremental validity does not exist to date. In the study by Lewis et al. (2002), the SIMS total score did not show incremental validity beyond MMPI-2 Fb, which was the best predictor of invalid status. Given the low PPP of the SIMS in this study, the authors reiterated the need to follow-up with more thorough assessment in cases of invalid SIMS scores. In the Edens et al.' study (2007), logistic regressions were used to examine whether the SIRS showed incremental validity over the SIMS; although the model was significant when the SIRS was added, predictive accuracy only increased from 69 to 72%. No other studies were found examining incremental validity of the SIMS.

On the other hand, if an evaluator wonders if the SIMS will improve their ability to rule out exaggeration/feigning beyond evaluator judgment based on record review and interview data only, then the SIMS does evince incremental validity (Dandachi-FitzGerald et al., 2017; Edens et al., 2007). In this context, the SIMS also possesses *incremental utility*. We define incremental utility as the extent to which a symptom validity test demonstrates incremental validity *and* provides value. Value is a subjective judgment that potentially involves several variables, but in relatively simple terms, a test has *value* if its benefits outweigh its costs (see Hunsley & Meyer, 2003, and Haynes & Lench, 2003, for detailed and erudite discussions of incremental validity and utility). If an evaluator wishes to conduct comprehensive exaggeration/feigning assessments only with evaluatees who show an elevated risk of exaggeration/feigning, then the SIMS has value as it takes only 15 min to administer, costs \$5.50 USD per test for administration and scoring (PAR, n.d.), and reliably rules out a sizeable portion of evaluatees from needing further assessment.

In conclusion, SIMS incremental validity is not well established when compared to other SVTs; however, this applies more so in the test battery use than in the screening use, and incremental validity arguably is less important in that situation. A small number of studies suggest incremental validity and utility for the SIMS compared to evaluator judgments about exaggeration/feigning based on interview and record review data alone.

Cut Scores and Hit Rates

A number of cutoff scores have been examined for the SIMS. The cutoff score published on the paper test forms and presented in the test manual is > 14 for the Total score (Widows & Smith, 2005). This cutoff recommendation was based on empirical validation of the SIMS described in the manual and using two samples of college students. Study participants were randomly assigned either to a control group of honest responders (HR) or to one of six experimental conditions, in which participants were asked to simulate psychosis (P), amnesia (A), depression (D), low intelligence (LI), neurologic impairment (NI), or to simulate “faking bad” (FB) (Widows & Smith, 2005, p. 12). The total sample was split into developmental ($n=238$) and cross-validation ($n=238$) subsamples. Using the developmental sample, the authors identified cutoff scores that optimally separated participants in the HR group from participants in the experimental conditions on each scale. Resulting cutoffs were as follows: Total > 14 , P > 1 , NI > 2 , AM > 2 , LI > 2 , and AF > 5 . Sensitivity for the total score cutoff was 94.63% with specificity of 87.88%. The authors concluded that respondents who obtain SIMS total score greater than 14 are in need of further evaluation due to endorsing a high number of atypical, improbable, inconsistent, or illogical symptoms.

In addition to the cutoff score of > 14 published in the manual, several additional cutoff scores have been examined in the extant literature (e.g., van Impelen et al., 2014). van Impelen et al. (2014) conducted a meta-analysis of studies examining SIMS cutoffs, separating published research into two categories: (1) known groups and (2) simulation studies. In the simulation studies, they reported that utilized cutoff scores were either > 14 or > 16 . They concluded that in those studies, SIMS scores differed significantly between experimental feigners and honest responders, and generally those studies reported rather high sensitivity values (0.63–1.00) with variable specificity (0.23–1.00). Effect sizes were variable as well, with Cohen's d values ranging from 0.5 to 4.7. In the known-groups studies, a number of cutoffs were reported > 13 , > 14 , > 16 , > 19 , > 21 , and > 23 . Sensitivity values in those studies were relatively low for cutoffs > 21 (0.68) and > 23 (0.55) but much higher (0.75–1.00) for lower cutoffs. Specificity values were 0.73 and 1.00 for cutoffs > 21 and > 23 , respectively, and they ranged between 0.37 and 0.93 for lower cutoffs. Overall, based on information presented by van Impelen et al. (2014), SIMS scores produce adequate sensitivity values at lower cutoffs, but specificity values are often low in known-group studies or highly variable in simulation studies. In summary, van Impelen et al. (2014, p. 1356) concluded that cutoffs > 14 and > 16 are “perilous” and that cutoff scores of > 19 (Clegg et al., 2009) or > 23 (Wisdom et al., 2010) should be considered.

In the present manuscript, we examined literature published to date (in English) that explored various SIMS cutoffs, and findings were organized based on the context in which the study was conducted: (1) research, (2) clinical, or (3) forensic settings. All studies included in the present review are outlined in Table 1. In a research setting, the initial validation study of the SIMS was conducted by Smith and Burger (1997). Participants in that study were college students ($N=476$) who were assigned to an honest responding group or one of seven simulation conditions (simulating psychosis, amnesic disorder, neurological impairment, mania, depression, low intelligence, and “faking bad”). Results of that study indicated that the SIMS total score had the highest sensitivity (95.6%) and overall efficiency (percent correctly classified as malingering or non-malingering; 94.5%) compared to other validity indices (such as the F and K scales of the MMPI and 16PF Faking Bad scale). Corresponding specificity (87.9%) was adequate. The authors concluded that their data supported the construct validity of the SIMS, and that individuals who obtain a total score > 14 on the SIMS may be “suspected of malingering” (p. 188) and may require a further evaluation. Another simulation study published approximately around the same time by Rogers et al. (1996) involved 53 adolescent offenders who were asked to complete the SIMS under honest and feigning conditions. The authors utilized a different SIMS total cutoff score of > 16 and concluded that this score was moderately effective as a screening measure in identifying feigned protocols. However, NPP rate was rather low (0.62). The authors examined other cutoff scores for the SIMS and concluded that a score of > 40 (PPP=0.49; NPP=0.94) would be more optimal in their sample of dually diagnosed adolescents who were undergoing court-referred residential psychological treatment. Similarly, Rogers et al. (2014) reported that at cutoff score of > 14 , specificity was very low (0.28) and concluded that a much higher SIMS total cutoff score (> 44) is needed to achieve high specificity (0.98) in their sample of mental health inpatients with extensive trauma histories who were asked to simulate disability.

In general, most simulation studies have examined two cutoff scores: > 14 and > 16 . Across all simulation studies included in the present review (Table 1), sensitivity values for cutoff > 14 ranged between 0.52 and 0.98, whereas specificity values ranged between 0.28 and 1.00 (although most studies indicated that this cutoff achieved specificity in the 0.79–1.00 range). For the cutoff score of > 16 , sensitivity values ranged between 0.36 and 0.98, whereas specificity values ranged between 0.23 and 1.00 (although most studies indicated that this cutoff score achieved specificity in the 0.83–1.00 range). Reported classification accuracy rates for a score of > 14 in simulation studies ranged between 0.57 and 0.95, whereas classification accuracy for > 16 ranged between 0.68 and 0.99 (see Table 1).

In clinical contexts, classification accuracy of the traditional cutoff score of > 14 produced mixed results. One of the first studies to raise concerns about the applicability of the traditional cutoff > 14 was published by Edens et al. (1999), who found that SIMS specificity rates were low in individuals reporting clinically significant distress and concluded that this cutoff score may be problematic due to potentially high false positive rates among clinical populations. Edens et al. (2007) later noted that at > 14 cutoff, specificity values ranged from 0.40 in psychiatric patients to 0.97 in controls, highlighting that although specificity was high in the control group, it was poor among patients with genuine psychiatric disorders. Some more recent studies have shown inadequate specificity (0.55) with higher sensitivity (0.76) for this cutoff (Benge et al., 2012), while other studies have revealed lower sensitivity (0.52–60) but moderate to high specificity values (0.75–0.98) associated with it (e.g., Harris & Merz, 2021; Puente-Lopez et al., 2020).

Regarding base rates of failure at different cutoff scores, research indicates that in some populations, failure rate at the cutoff > 14 may be rather high. For example, Ord et al. (2021) reported failure rate of approximately 45% for this cutoff in a sample of Iraq and Afghanistan combat veterans, compared to 14% failure rate when > 23 cutoff was utilized, suggesting that higher cutoffs may be more appropriate for use in certain populations. This high rate of overreporting may be attributed to the fact that combat veterans often have multiple comorbid conditions and subsequently may display higher levels of genuine psychopathology, which may in turn elevate their SIMS scores. These authors suggested that the traditional SIMS cutoff (> 14) may not be appropriate for use in combat-exposed veterans and that further research is needed to examine the use of SIMS in this population. Similarly, Harris and Merz (2021) examined different SIMS cutoffs in a sample of 110 patients at an adult neuropsychology clinic and found a high rate of elevations at the standard cutoff score of > 14 (45.5%), which corresponded to a failure rate of 24.4% in a group of “low suspicion” cases and a failure rate of 95.7% in “high suspicion” cases. They concluded that SIMS scores should be interpreted with caution and the score of > 14 could be used for screening purposes to determine whether a more thorough follow-up would be needed. They recommended a cutoff score of > 16 to be used in neuropsychological populations but noted that it still had only modest specificity.

Because the traditional cutoff score of > 14 may not be optimal for the use in clinical populations, a number of alternative cutoffs have been examined. For example, Dandachi-FitzGerald et al. (2020) examined SIMS cutoffs > 16 and > 19 in 120 clinical patients diagnosed with dementia, mild cognitive impairment, or Parkinson’s disease. They reported that only 10% obtained a score > 19 , and 12.5% obtained a score > 16 , concluding that the SIMS may have

adequate classification accuracy rates at these cutoffs for these populations. Another study (Czornik et al., 2021) examined a sample of 54 memory clinic outpatients and found the overreporting rate at the cutoff > 16 was 14.8%, and elevating the cutoff score to > 21 resulted in a 7.5% rate. Bengue et al. (2012) explored the diagnostic utility of the SIMS in identifying psychogenic non-epileptic events and examined a number of cutoff scores ranging from 10 to 40. Cutoff scores > 14 and > 16 resulted in rather high sensitivity numbers (0.76 and 0.71, respectively), but corresponding specificity values were low (0.55 and 0.69). The first cutoff score to achieve an adequate specificity value of 0.90 was > 23 , but with an accompanying sizable drop in sensitivity (0.39). Finally, Peters et al. (2013) examined 41 patients diagnosed with schizophrenia and 43 non-clinical controls. They reported that in their schizophrenia sample, the specificity rate at the cutoff > 16 was 0.71, and that it increased to 0.81 when the cutoff score was raised to > 19 . To achieve a specificity rate of 0.90 or above, the cutoff > 21 would need to be utilized in their sample. Taken as a whole, findings of studies conducted in clinical contexts indicate that the traditional cutoff > 14 should be used with caution, even for screening purposes, and that higher cutoffs may be more appropriate for various clinical populations.

In forensic settings, a traditionally utilized cutoff > 14 demonstrated high sensitivity (0.85–1.00) values, which were similar to cutoffs > 13 (0.87–0.96) and > 16 (0.75–1.00) (Table 1). However, these cutoffs corresponded to rather low specificity values in forensic settings (0.37–0.67) (Table 1). To achieve acceptable specificity (≥ 0.90), the cutoff score needed to be increased to > 22 (Wisdom et al., 2010), whereas cutoff > 23 resulted in perfect specificity of 1.00 (Wisdom et al., 2010). Similarly, Clegg et al. (2009) noted that increasing the SIMS cutoff score to > 19 may improve the utility of this measure in the assessment of symptom validity among disability claimants. These authors examined 56 individuals seeking disability classified into honest and suspected malingerers based on their SIRS scores, and 60 individuals from the community who completed SIMS honestly or as if they were malingering depression. They found that individuals in both malingering groups received significantly higher SIMS total scores compared to the honest groups. However, no significant differences were found between the scores of two malingering groups. The authors concluded that cutoff scores > 14 and > 16 had excellent sensitivity but low specificity; therefore, increasing the SIMS total cutoff score to at least > 19 may improve the utility of this measure in the population of Social Security Disability claimants. Similarly, other authors have suggested that cutoff > 16 may be inadequate for forensic populations. For example, Lewis, et al. (2002) examined 64 men undergoing pretrial forensic assessments and reported 1.00 sensitivity

but only 0.61 specificity rates for the cutoff > 16 . Alwes et al. (2008) also reported moderate to high sensitivity rates (0.75–0.96) but low specificity rates (0.60–0.67) for the cutoff > 16 in their sample of 308 individuals who completed a neuropsychiatric evaluation for workers' compensation or personal injury claims. In summary, studies in forensic settings have shown that the traditional cutoff > 14 and the cutoff > 16 produced inadequate specificity. To achieve acceptable specificity (≥ 0.90) in forensic populations, higher cutoff scores appear to be generally more appropriate.

Cutoff Scores for SIMS Subscales

As stated earlier, the following cutoff scores were identified in the SIMS manual for each subscale: $P > 1$, $NI > 2$, $AM > 2$, $LI > 2$, and $AF > 5$ (Widows & Smith, 2005). Sensitivity values of subscales ranged from 0.76 (AF) to 0.89 (AM) at these cutoffs. Specificity values for the subscales ranged from 0.76 (P) to 0.9 (NI and AF). Smith and Burger (1997) reported the following sensitivity values for the subscales at the published cutoffs: $AF = 0.75$, $P = 0.80$, $N = 0.85$, $LI = 0.85$, $AM = 0.88$. Corresponding specificity values were reported as follows: $AF = 0.73$, $P = 0.73$, $N = 0.76$, $LI = 0.52$, $AM = 0.91$. Bengue et al. (2012) examined various cutoffs for the NI and AF scales specifically. They reported that the cutoff score > 2 for the NI scale was associated with the sensitivity value of 0.75 and specificity value of only 0.62 in their sample of inpatients with epilepsy or psychogenic non-epileptic events. The traditionally utilized cutoff > 5 for the AF subscale corresponded with sensitivity of 0.58 and specificity of 0.72. Adequate specificity (> 0.90) was achieved at the cutoff > 5 for NI and the cutoff > 7 for AF, indicating that these scores may be more appropriate for use in populations with neurological concerns. Parks et al. (2017) reported the following sensitivity values for each subscale at the published cutoffs in a sample of 78 undergraduate simulators: $P = 0.35$, $LI = 0.42$, $AM = 0.64$, $NI = 0.67$, $AF = 0.92$. Specificity values were not reported. Finally, Harris and Merz (2021) reported rather high failure rates for most subscales based on the published cutoffs in their sample of 110 adult neuropsychology outpatients: $LI = 36.4\%$, $AF = 47.3\%$, $NI = 55.5\%$, and $AM = 64.5\%$. The only exception was P subscale (feigning rate was 17.3%), which was more consistent with the expected rate in clinical populations. In summary, although most SIMS subscales displayed moderate to high sensitivity, not all of them achieved adequate specificity values at the cutoff scores published in the manual, based on results of available empirical studies.

In addition to the five subscales identified in the SIMS manual (Widows & Smith, 2005), some authors have proposed additional subscales or indices. For example, Rogers et al. (2005) developed a new index: an arithmetic difference

between scores on AF and N subscales. They suggested a cutoff > 0 for that index, which corresponded to sensitivity of 1.00 but low specificity of 0.31 (total hit rate = 0.77). They found significant differences on that new index between the malingering condition and both factitious conditions in a sample of 65 psychology doctoral students using a simulation study design.

Further, the previously mentioned RS and SC subscales (Rogers et al., 2014) produced high specificity (0.98) with moderate sensitivity values of 0.42 (RS) and 0.67 (SC) at the recommended cutoff > 6 . The authors concluded that these scales had potential utility in identifying simulated mental disorders. Later, Edens et al. (2020) examined classification accuracy of the RS and SC subscales in three archival samples: (1) 115 inmates (general population and inmates in a prison psychiatric unit receiving treatment for mental disorder), (2) 196 college students, and (3) 48 community-dwelling adults. They concluded that the suggested cutoff score > 6 for both subscales produced relatively low sensitivity values (ranging from 0 to 0.56 for different subsamples), although with corresponding high specificity (ranging from 0.87 to 1.00). These scales have been criticized (Cernovsky & Ferrari, 2020), and further validation is warranted before employing these scales clinically.

Interpretive Recommendations

Based on available known groups studies, we offer the following interpretive recommendations. These are general guidelines to consider. First, SIMS cutoff scores should be determined *a priori*, after taking into account: the evaluative setting's feigning base rate (as best it can be estimated), the extent to which (in a given evaluative setting) examinees with genuine psychopathology are likely to produce elevated SIMS scores, and one's willingness and ability to conduct a comprehensive feigning assessment when respondents produce SIMS scores above the pre-established cutoff score. Second, in most contexts, the manual cutoff score of > 14 will result in a high number of false positives and should generally be avoided. Therefore, we recommend using a cutoff score of > 19 for screening purposes and > 23 when used as convergent data (i.e., in a test battery with multiple SVTs). Alternately, in either a screening or battery use, SIMS users may also consider a graded approach for interpreting SIMS scores as follows: possibly invalid > 16 , probably invalid > 19 , very likely invalid > 23 . Of note, we recommend the use of terminology other than malingering when describing test scores, given that the SIMS does not measure malingering per se, but instead measures feigning. Regardless of which interpretive approach is used, a SIMS Total cutoff score of at least > 23 is recommended for forensic situations.

Beyond the total score, caution is warranted when considering elevated subscales, given lack of research on subscale classification accuracy. Invalid subscale scores based on manual cutoffs might be best for explaining what types of exaggerated symptoms are driving an elevated Total score, rather than being used in isolation or when the total score is not elevated. Finally, when administering the SIMS to groups with certain characteristics, such as undergraduate students, or patients presenting with PTSD symptoms, we recommend consulting the research literature for studies examining similar populations. For example, in some research settings using the SIMS to exclude invalid subjects in a sample of healthy undergraduates, using the manual cutoff score of > 14 might be appropriate. In other contexts, more extreme cutoff scores might also be relevant (e.g., > 44 for psychiatric inpatients; Rogers et al., 2014). Table 1 can help readers identify studies with samples relevant to their own settings.

Strengths and Weaknesses

Strengths

The SIMS has a number of strengths that likely contribute to its popularity. First, the SIMS offers evaluators an efficient and effective method to rule out feigning among a sizeable proportion of examinees, so that they can devote the additional resources necessary for comprehensive symptom validity assessment to those demonstrating a greater likelihood of feigning. Second, the SIMS has several practical advantages: it takes only about 15 min to administer, is easy to score, is written at a fifth-grade reading level, costs less per test than many other symptom validity measures, and is available in a variety of administration modalities and several languages. Additionally, the SIMS shows incremental validity when compared to clinical judgments based on interview and record review data alone. Finally, the SIMS has been validated with a variety of cultural groups in the USA and Europe.

Weaknesses

Despite the noted strengths of the SIMS, several limitations emerge from existing research. First, the SIMS does not appear to show incremental validity beyond the validity scales of standard multiscale inventories like the MMPI-2-RF or the PAI. However, this limitation highlights the instrument's two primary uses: (1) as a screener for identifying situations where more thorough symptom validity assessment is needed and (2) as a supplement in a battery of validity measures. Second, the SIMS does not exhibit robust

discriminant validity when evaluating groups with severe psychopathology, and evaluators should use higher cutoff scores more generally (e.g., total > 19 [van Impelen, et al., 2014] or > 23 [Wisdom, et al., 2010]) or based on studies using samples with similar conditions (see Table 1).

The most notable limitation to the SIMS is that although SIMS items are described as containing “implausible, rare, atypical, or extreme symptoms that bona fide patients tend not to present” (Mazza et al., 2019, p. 5), this does not appear to be the case for all items. For example, two items describe trouble sleeping in general and middle insomnia specifically, neither of which are worded as extreme or severe. Yet, sleep disruption is a common complaint, with the *DSM-5* noting that about a third of the general population suffers from insomnia symptoms (American Psychiatric Association, 2013). Similarly, from the content perspective, many items, particularly on the AF and AM scales, reflect genuine complaints common in clinical populations. In other words, many SIMS items represent quasi-rare symptoms (Rogers, 2018), as opposed to bizarre ones. A final weakness is the use of the term malingering in the SIMS name. Because the SIMS does not actually measure malingering, the test should be referred to as a symptom validity test, not a malingering test.⁵

One particular group of authors has recently focused on criticizing the SIMS to the extent of calling the use of the SIMS malpractice (Cernovsky & Diamond, 2020; Cernovsky & Fattahi, 2020; Cernovsky et al., 2019a, 2019b, 2019c, 2020) and claiming that the SIMS is “essentially an iatrogenic pseudoscientific instrument” (Cernovsky et al., 2020, p. 9). They make a few valid points, notably that there is an issue of false positives using the manual cutoff scores, that many individual items in the SIMS can reflect genuine/common problems, and that additional validation in true medical groups is needed. Nevertheless, the group’s opinion appears to be biased, and their theses tend to be problematic across a few themes. First, the authors show a misunderstanding that the SIMS is intended to be used in isolation to diagnose malingering specifically. They also describe professionals misusing the test in this manner but blame the test itself. The authors do not cite the numerous instances in the SIMS manual where this is specifically not the recommended use (it is recommended to be used either as a screener or as one part of a multi-method full battery evaluation). Thus, individuals using the SIMS alone to diagnose malingering are not using the test correctly, which is not an inherent flaw or weakness of the test itself. Second, there is an apparent view that concussion patients (specifically, civil forensic MVA patients)

should be expected to show chronic impairments, which has not been supported in the vast concussion literature, with persisting symptoms often best explained by non-injury factors, most notably litigation (Carroll et al., 2004, 2014; Cassidy et al., 2014; McCrea, 2008).

Third, the authors appear to lack an understanding of symptom exaggeration as opposed to symptom fabrication: i.e., an SVT can be made of “real” (non-bizarre) symptoms, but endorsing a high number can indicate exaggeration. Many SVTs do in fact effectively use real symptoms, where the high endorsement of symptoms is indicative of symptom exaggeration (e.g., the *F* scale on the MMPI). Rogers (2018) refers to these detection strategies as quasi-rare symptoms among a variety of amplified detection strategies. Last, the authors engage in the relying on the error of psychopathology as superordinate (Merten & Merckelbach, 2013), or explaining feigning behaviors using psychiatric symptoms (e.g., “cry for help”). In short, readers would be cautioned against using these papers as the basis of evaluating the SIMS.

Future Perspectives

There are several opportunities for future research to further the utility of the SIMS. More generally, there is room for additional studies using the SIMS across a variety of contexts (research, clinical, forensic), psychological and medical conditions (e.g., PTSD, TBI, neurological conditions), and populations, particularly military and veteran samples. SIMS research would benefit from additional convergent validity studies. Further research on the original five subscales would also be welcome given the majority of existing research has focused on the Total score. Also, an update to the manual might be helpful, incorporating the numerous studies reviewed here, and in the van Impelen et al. (2014) meta-analysis, into updated recommendations for cutoff scores across a variety of different situations. Although several translations have been created, additional translations and further validation in diverse cultures might also be considered for future work.

The SIMS manual (Widows & Smith, 2005) describes the measure as a screening instrument, but additional research has evaluated the utility of higher cutoff scores in some situations. For example, van Impelen et al.’s meta-analysis (2014) suggests using lower cutoff scores (> 14 or > 16) for most screening purposes and higher cutoff scores (> 19 or > 24) when using the SIMS as part of a test battery. Both the manual and the aforementioned meta-analysis note that the SIMS is also useful when added to other SVTs for the purpose of providing convergent data, and future studies might further explore how to best use the SIMS when employed in a battery using numerous validity indicators (e.g., incorporating interpretation of multivariate base rates).

⁵ It should be noted that the SIMS manual refers to the Slick et al. (1999) criteria and best practices of using multiple sources of data when evaluating malingering.

In addition to screening versus test battery cutoff scores, future research could further assess the idea that the SIMS measures two separate feigning constructs: symptom magnification and symptom fabrication. This classification has been explored in several studies using the SIMS (Mazza et al., 2019; Orrù et al., 2021), and additional validation of these underlying constructs could lead to improved precision when describing and interpreting feigning behavior. Similarly, the several short forms (Malcore et al., 2015; Orrù et al., 2021; Spencer et al., 2021) and different subscales (Rogers et al., 2014) need cross-validation before being deployed clinically. Relatedly, there are few studies that have evaluated SIMS scores in relation to performance validity tests, and similar to the RBS scale on the MMPI-2-RF/3, the AM SIMS subscale in particular might be useful for screening for invalid performance, as was found by Spencer et al. (2021). Finally, as additional studies are published on the SIMS, an updated meta-analysis might prove beneficial.

Funding This work was supported by the Salisbury VA Health Care System, Mid-Atlantic (VISN 6) Mental Illness Research, Education, and Clinical Center (MIRECC) and the Department of Veterans Affairs Office of Academic Affiliations Advanced Program in Mental Illness, Research, and Treatment.

Declarations

Conflict of Interest The authors declare no competing interests.

Disclaimer The views, opinions, and findings contained in this article are those of the authors and should not be construed as an official Veterans Affairs position, policy, or decision, unless so designated by other official documentation.

References

- Alwes, Y. R., Clark, J. A., Berry, D. T. R., & Granacher, R. P. (2008). Screening for feigning in a civil forensic setting. *Journal of Clinical and Experimental Neuropsychology*, 30, 133–140.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). <https://doi.org/10.1176/appi.books.9780890425596>
- Ardic, F. C., Kose, S., Solmaz, M., Kulacaoglu, F., & Balcioglu, Y. H. (2019). Reliability, validity, and factorial structure of the Turkish version of the Structured Inventory of Malingered Symptomatology (Turkish SIMS). *Psychiatry and Clinical Psychopharmacology*, 29(2), 182–188. <https://doi.org/10.1080/24750573.2019.1599237>
- Benge, J. F., Wisdom, N. M., Collins, R. L., Franks, R., LeMaire, A., & Chen, D. K. (2012). Diagnostic utility of the Structured Inventory of Malingered Symptomatology for identifying psychogenic non-epileptic events. *Epilepsy & Behavior*, 24(4), 439–444. <https://doi.org/10.1016/j.yebeh.2012.05.007>
- 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(4), 404–417. <https://doi.org/10.1016/j.spinee.2004.11.016>
- Boskovic, I., van der Heide, D., Hope, L., Merckelbach, H., & Jelicic, M. (2017). Plausibility judgments of atypical symptoms across cultures: An explorative study among Western and non-Western experts. *Psychological Injury and Law*, 10, 274–281. <https://doi.org/10.1007/s12207-017-9294-6>
- Carroll, L. J., Cassidy, J. D., Cancelliere, C., Côté, P., Hincapié, C. A., Kristman, V. L., Holm, L. W., Borg, J., Nygren-de Boussard, C., & Hartvigsen, J. (2014). Systematic review of the prognosis after mild traumatic brain injury in adults: Cognitive, psychiatric, and mortality outcomes: Results of the International Collaboration on Mild Traumatic Brain Injury Prognosis. *Archives of Physical Medicine and Rehabilitation*, 95(3 Suppl), S152–S173. <https://doi.org/10.1016/j.apmr.2013.08.300>
- Carroll, L. J., Cassidy, J. D., Peloso, P. M., Borg, J., von Holst, H., Holm, L., Paniak, C., Pépin, M., & WHO Collaborating Centre Task Force on Mild Traumatic Brain Injury (2004). Prognosis for mild traumatic brain injury: Results of the WHO Collaborating Centre Task Force on Mild Traumatic Brain Injury. *Journal of rehabilitation medicine*, (43 Suppl), 84–105. <https://doi.org/10.1080/16501960410023859>
- Cassidy, J. D., Cancelliere, C., Carroll, L. J., Côté, P., Hincapié, C. A., Holm, L. W., Hartvigsen, J., Donovan, J., Nygren-de Boussard, C., Kristman, V. L., & Borg, J. (2014). Systematic review of self-reported prognosis in adults after mild traumatic brain injury: Results of the International Collaboration on Mild Traumatic Brain Injury Prognosis. *Archives of Physical Medicine and Rehabilitation*, 95(3 Suppl), S132–S151. <https://doi.org/10.1016/j.apmr.2013.08.299>
- Cernovsky, Z. Z., & Diamond, D. M. (2020). High risk of false classification of injured people as malingerers by the Structured Inventory of Malingered Symptomatology (SIMS): A review. *Archives of Psychiatry and Behavioral Sciences*, 3(2), 30–38.
- Cernovsky, Z. Z., & Ferrari, J. R. (2020). Rogers's RS and SC malingering scales derived from the SIMS. *Archives of Psychiatry and Behavioral Sciences*, 3(1), 34–44.
- Cernovsky, Z. Z., Ferrari, J. R., & Mendonça, J. D. (2019a). Deceptive clinical diagnosing of malingering via Structured Inventory of Malingered Symptomatology. *Archives of Psychiatry and Behavioral Sciences*, 2(1), 44–49.
- Cernovsky, Z. Z., Ferrari, J. R., & Mendonça, J. D. (2019b). Pseudodiagnoses of malingering of neuropsychological symptoms in survivors of car accidents by the Structured Inventory of Malingered Symptomatology. *Archives of Psychiatry and Behavioral Sciences*, 2(1), 55–65.
- Cernovsky, Z., & Fattahi, M. (2020). Meta-analysis of scale cutoffs in the Structured Inventory of Malingered Symptomatology. *International Journal of Psychological Sciences*, 1(1), 1–7.
- Cernovsky, Z. Z., Mendonça, J. D., Bureau, Y. R. J., & Ferrari, J. R. (2019c). Criterion validity of low intelligence scale of the SIMS. *International Journal of Psychology Sciences*, 1(1), 3–5.
- Cernovsky, Z. Z., Mendonça, J. D., & Ferrari, J. R. (2020). Meta-analysis of SIMS scores of survivors of car accidents and of instructed malingerers. *Archives of Psychiatry and Behavioral Sciences*, 3(1), 1–11.
- 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. *Journal of Forensic Psychiatry & Psychology*, 20(2), 239–254. <https://doi.org/10.1080/14789940802267760>
- Czornik, M., Merten, T., & Lechner, J. (2021). Symptom and performance validation in patients with subjective cognitive decline and mild cognitive impairment. *Applied Neuropsychology: Adult*, 28(3), 269–281. <https://doi.org/10.1080/23279095.2019.1628761>
- Dandachi-FitzGerald, B., Duits, A. A., Leentjens, A. F., Verhey, F. R., & Ponds, R. W. (2020). Performance and symptom validity

- assessment in patients with apathy and cognitive impairment. *Journal of the International Neuropsychological Society*, 26(3), 314–321. <https://doi.org/10.1017/S1355617719001139>
- Dandachi-FitzGerald, B., & Merckelbach, H. (2013). Feigning ≠ feigning a memory disorder: The Medical Symptom Validity Test as an example. *Journal of Experimental Psychopathology*, 4, 46–63. <https://doi.org/10.5127/jep.025511>
- Dandachi-FitzGerald, B., Merckelbach, H., & Ponds, R. W. (2017). Neuropsychologists' ability to predict distorted symptom presentation. *Journal of Clinical and Experimental Neuropsychology*, 39(3), 257–264. <https://doi.org/10.1080/13803395.2016.1223278>
- Dandachi-FitzGerald, B., Ponds, R. W. H. M., & Merten, T. (2013). Symptom validity and neuropsychological assessment: A survey of practices and beliefs of neuropsychologists in six European countries. *Archives of Clinical Neuropsychology*, 28(8), 771–783. <https://doi.org/10.1093/arclin/act073>
- Dandachi-FitzGerald, B., Ponds, R. W. H. M., Peters, M. J. V., & Merckelbach, H. (2011). Cognitive underperformance and symptom over-reporting in a mixed psychiatric sample. *The Clinical Neuropsychologist*, 25(5), 812–828. <https://doi.org/10.1080/13854046.2011.583280>
- Edens, J. F., Otto, R. K., & Dwyer, T. (1999). Utility of the Structured Inventory of Malingered Symptomatology in identifying persons motivated to malingering psychopathology. *Journal of the American Academy of Psychiatry and 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(1), 33–42.
- Edens, J. F., Truong, T. N., & Otto, R. K. (2020). Classification accuracy of the rare symptoms and symptom combinations scales of the Structured Inventory of Malingered Symptomatology in three archival samples. *Law and Human Behavior*, 44(2), 167–177. <https://doi.org/10.1037/lhb0000361>
- Franklin, C., Repasky, S., Thompson, K., Shelton, S., & Uddo, M. (2002). Differentiating overreporting and extreme distress: MMPI-2 use with compensation-seeking veterans with PTSD. *Journal of Personality Assessment*, 79(2), 274–285.
- Freeman, T., Powell, M., & Kimbrell, T. (2008). Measuring symptom exaggeration in veterans with chronic posttraumatic stress disorder. *Psychiatry Research*, 158(3), 374–380.
- Frueth, B. C., Hamner, M. B., Cahill, S. P., Gold, P. B., & Hamlin, K. L. (2000). Apparent symptom overreporting in combat veterans evaluated for PTSD. *Clinical Psychology Review*, 20(7), 853–885.
- González Ordi, H., Capilla Ramírez, P., Santamaría Fernández, P., & Casado Morales, M. I. (2012). Abordaje multidisciplinar para la detección de la simulación en lumbalgia crónica [A multidisciplinary approach to the detection of malingering in chronic lower back pain]. *Trauma*, 23, 145–154.
- Goodwin, B. E., Sellbom, M., & Arbisí, P. A. (2013). Posttraumatic stress disorder in veterans: The utility of the MMPI-2-RF validity scales in detecting overreported symptoms. *Psychological Assessment*, 25(3), 671–678. <https://doi.org/10.1037/a0032214>
- Giger, P., Merten, T., Merckelbach, H., & Oswald, M. (2010). Detection of feigned crime-related amnesia: A multi-method approach. *Journal of Forensic Psychology Practice*, 10, 440–463.
- Giomini, L., Viglione, D. J., Pignolo, C., & Zennaro, A. (2018). A clinical comparison, simulation study testing the validity of SIMS and IOP-29 with an Italian sample. *Psychological Injury and Law*, 11(4), 340–350. <https://doi.org/10.1007/s12207-018-9314-1>
- Grant, A. F., Lace, J. W., Teague, C. L., Lowell, K. T., Ruppert, P. D., Garner, A. A., & Gfeller, J. D. (2020). Detecting feigned symptoms of depression, anxiety, and ADHD, in college students with the structured inventory of malingered symptomatology. *Applied Neuropsychology: Adult*. Advance online publication. <https://doi.org/10.1080/23279095.2020.1769097>
- Graue, L. O., Berry, D. T. R., Clark, J. A., Sollman, M. J., Cardi, M., Hopkins, J., & Werline, D. (2007). Identification of feigned mental retardation using the new generation of malingered detection instruments: Preliminary findings. *The Clinical Neuropsychologist*, 21, 929–942.
- Harris, M., & Merz, Z. C. (2021). High elevation rates of the Structured Inventory of Malingered Symptomatology (SIMS) in neuropsychological patients. *Applied Neuropsychology: Adult*. Advance online publication. <https://doi.org/10.1080/23279095.2021.1875227>
- Haynes, S. N., & Lench, H. C. (2003). Incremental validity of new clinical assessment measures. *Psychological Assessment*, 15(4), 456–466. <https://doi.org/10.1037/1040-3590.15.4.456>
- 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*, 1(4), 23–52.
- Hunsley, J., & Meyer, G. (2003). The incremental validity of psychological testing and assessment: Conceptual, methodological, and statistical issues. *Psychological Assessment*, 15(4), 446–455.
- Jelicic, M., Ceunen, E., Peters, M. J. V., & Merckelbach, H. (2011). Detecting coached feigning using the Test of Memory Malingering (TOMM) and the Structured Inventory of Malingered Symptomatology (SIMS). *Journal of Clinical Psychology*, 67, 850–855.
- 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 Behavioral Assessment*, 28(1), 19–22.
- Jelicic, M., Merckelbach, H., Candel, I., & Geraerts, E. (2007a). Detection of feigned cognitive dysfunction using special malingering tests: A simulation study in naïve and coached malingerers. *International Journal of Neuroscience*, 117, 1185–1192.
- Jelicic, M., Peters, M. J. V., Leckie, V., & Merckelbach, H. (2007b). Basic knowledge of psychopathology does not undermine the efficacy of the Structured Inventory of Malingered Psychopathology (SIMS) to detect feigned psychosis. *Netherlands Journal of Psychology*, 63, 107–110.
- Jelicic, M., van Gaal, M., & Peters, M. J. V. (2013). Expert knowledge doesn't help: Detecting feigned psychosis in people with psychiatric expertise using the Structured Inventory of Malingered Symptomatology (SIMS). *Journal of Experimental Psychopathology*, 4, 38–77.
- Kelley, T. L. (1927). *Interpretation of educational measurements*. New York: Macmillan.
- Larrabee, G. L. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist*, 17, 410–425. <https://doi.org/10.1076/clin.17.3.410.18089>
- 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.
- Malcore, S. A., Schutte, C., Dyke, S. A., & Axelrod, B. N. (2015). The development of a reduced-item Structured Inventory of Malingered Symptomatology (SIMS). *Psychological Injury and Law*, 8(2), 95–99. <https://doi.org/10.1007/s12207-015-9214-6>
- Martin, P. K., Schroeder, R. W., & Odland, A. P. (2015). Neuropsychologists' validity testing beliefs and practices: A survey of North American professionals. *The Clinical Neuropsychologist*, 29(6), 741–776. <https://doi.org/10.1080/13854046.2015.1087597>
- Martin, P. K., & Schroeder, R. W. (2020). Base rates of invalid test performance across clinical non-forensic contexts and settings. *Archives of Clinical Neuropsychology*, acaa107. <https://doi.org/10.1093/arclin/aca017>
- Mazza, C., Orrù, G., Burla, F., Monaro, M., Ferracuti, S., Colasanti, M., & Roma, P. (2019). Indicators to distinguish symptom accentuators from symptom producers in individuals with a diagnosed adjustment

- disorder: A pilot study on inconsistency subtypes using SIMS and MMPI-2-RF. *PLoS One*, 14(12), e0227113. <https://doi.org/10.1371/journal.pone.0227113>
- McCrea, M. A. (2008). *Mild traumatic brain injury and postconcussion syndrome: The new evidence base for diagnosis and treatment*. Oxford University Press.
- Merckelbach, H., Prins, C., Boskovic, I., Niesten, I., & à Campo, J. (2018). Alexithymia as a potential source of symptom over-reporting: An exploratory study in forensic patients and non-forensic participants. *Scandinavian Journal of Psychology*, 59(2), 192–197. <https://doi.org/10.1111/sjop.12427>
- Merckelbach, H. & Collaris, J. (2012). Mother Teresa doesn't help here: Lack of moral priming effects on malingered symptom reports and what we can learn from it. *Psychologica Belgica*, 52(2–3), 271–285.
- 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.
- Merten, T., Merckelbach, H. (2013). Symptom validity testing in somatoform and dissociative disorders: A critical review. *Psychological Injury and Law*, 6, 122–137. <https://doi.org/10.1007/s12207-013-9155-x>
- Merten, T., Kaminski, A., & Pfeiffer, W. (2020). Prevalence of overreporting on symptom validity tests in a large sample of psychosomatic rehabilitation inpatients. *The Clinical Neuropsychologist*, 34(5), 1004–1024.
- Morey, L. C. (2007). The Personality Assessment Inventory professional manual. In F. L., Lutz (Ed.), *Psychological Assessment Resources*.
- Nijdam-Jones, A., & Rosenfeld, B. (2017). Cross-cultural feigning assessment: A systematic review of feigning instruments used with linguistically, ethnically, and culturally diverse samples. *Psychological Assessment*, 29(11), 1321–1336. <https://doi.org/10.1037/pas0000438>
- Ord, A. S., Shura, R. D., Sansone, A. R., Martindale, S. L., Taber, K. H., & Rowland, J. A. (2021). Performance validity and symptom validity tests: Are they measuring different constructs? *Neuropsychology*. Advance online publication. <https://doi.org/10.1037/neu0000722>
- Orrù, G., Mazza, C., Monaro, M., Ferracuti, S., Sartori, G., & Roma, P. (2021). The development of a short version of the SIMS using machine learning to detect feigning in forensic assessment. *Psychological Injury and Law*, 14(1), 46–57. <https://doi.org/10.1007/s12207-020-09389-4>
- Oudman, E., Krooshof, E., van Oort, R., Lloyd, B., Wijnia, J. W., & Postma, A. (2020). Effects of Korsakoff Amnesia on performance and symptom validity testing. *Applied Neuropsychology: Adult*, 27(6), 549–557. <https://doi.org/10.1080/23279095.2019.1576180>
- Oxford University Press. (2018). Maligner. In *Oxford English Dictionary* (3rd ed., rev.). Retrieved August 18, 2021, from <https://www.oed.com/view/Entry/112944>
- Parks, A. C., Gfeller, J., Emmert, N., & Lammert, H. (2017). Detecting feigned postconcussional and posttraumatic stress symptoms with the structured inventory of malingered symptomatology (SIMS). *Applied Neuropsychology: Adult*, 24(5), 429–438.
- Peters, M. J. V., et al. (2013). Assessing the boundaries of symptom over-reporting using the Structured Inventory of Malingered Symptomatology in a clinical schizophrenia sample: Its relation to symptomatology and neurocognitive dysfunctions. *Journal of Experimental Psychopathology*, 4(1), 64–77.
- Puente-López, E., Pina, D., Ruiz-Hernández, & Llor-Esteban, B. (2020). Diagnostic accuracy of the Structured Inventory of Malingered Symptomatology (SIMS) in motor vehicle accident patients. *The Journal of Forensic Psychiatry & Psychology*, 1 – 24.
- Rogers, R. (2018). An introduction to response styles. In R. Rogers & S. D. Bender (Eds.), *Clinical assessment of malingering and deceptions* (4th ed., pp. 3–17). Guilford Press.
- 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., Hinds, J. D., & Sewell, K. W. (1996). Feigning psychopathology among adolescent offenders: Validation of the SIRS, MMPI-A, and SIMS. *Journal of Personality Assessment*, 67(2), 244–257. https://doi.org/10.1207/s15327752jpa6702_2
- Rogers, R., Jackson, R. L., & Kaminski, P. L. (2005). Factitious psychological disorders: The overlooked response style in forensic evaluations. *Journal of Forensic Psychology Practice*, 5(1), 21–41. https://doi.org/10.1300/J158v05n01_02
- Rogers, R., Robinson, E. V., & Gillard, N. D. (2014). The SIMS screen for feigned mental disorders: The development of detection-based scales. *Behavioral Sciences & the Law*, 32(4), 455–466. <https://doi.org/10.1002/bsl.2131>
- Slick, D. J., Sherman, E. M., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, 13(4), 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(2), 183–189. <http://jaapl.org/content/jaapl/25/2/183.full.pdf>
- Spencer, R. J., Gradwohl, B. D., & Kordovski, V. M. (2021). Initial validation of short forms of the SIMS for neuropsychological evaluations. *Psychological Injury and Law*, 14(1), 37–45. <https://doi.org/10.1007/s12207-020-09394-7>
- van der Heide, D., & Merckelbach, H. (2016). Validity of symptom reports of asylum seekers in a psychiatric hospital: A descriptive study. *International Journal of Law and Psychiatry*, 49 (A), 40–46. <https://doi.org/10.1016/j.ijlp.2016.05.007>
- van Impelen, A., Merckelbach, H., Jellic, M., & Merten, T. (2014). The Structured Inventory of Malingered Symptomatology (SIMS): A systematic review and meta-analysis. *The Clinical Neuropsychologist*, 28(8), 1336–1365. <https://doi.org/10.1080/13854046.2014.984763>
- 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.
- Widows, M. R., & Smith, G. P. (2005). *SIMS: Structured Inventory of Malingered Symptomatology: Professional manual*. PAR.
- Wisdom, N. M., Callahan, J. L., & Shaw, T. G. (2010). Diagnostic utility of the Structured Inventory of Malingered Symptomatology to detect malingering in a forensic sample. *Archives of Clinical Neuropsychology*, 25(2), 118–125.
- Wolf, E. J., Ellickson-Larew, S., Guetta, R. E., Escarfulleri, S., Ryabchenko, K., & Miller, M. W. (2020). Psychometric performance of the Miller Forensic Assessment of Symptoms Test (M-FAST) in veteran PTSD assessment. *Psychological Injury and Law*, 13(3), 284–302.
- World Health Organization (2021). *International Classification of Diseases 11th Revision, ICD-11 for Mortality and Morbidity Statistics*. Version: 05–2021. QC30 Malingering. <https://icd.who.int/browse/11/l-m/en#/http://id.who.int/icd/entity/1136473465>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.