The Structured Inventory of Malingered Symptomatology (SIMS): A Systematic Review and Meta-Analysis

Alfons van Impelen[a], Harald Merckelbach[a], Marko Jelicic[a] &
Thomas Merten[b]

[a] Forensic Psychology Section, Maastricht University, Maastricht,
The Netherlands

[b] Department of Neurology, Vivantes Klinikum im Friedrichshain,
Berlin, Germany
Published online: 10 Dec 2014.

Click for updates

PLEASE SCROLL DOWN FOR ARTICLE

The *Structured Inventory of Malingered Symptomatology* (SIMS): A Systematic Review and Meta-Analysis

Alfons van Impelen[1], Harald Merckelbach[1], Marko Jelicic[1], and Thomas Merten[2]

[1]Forensic Psychology Section, Maastricht University, Maastricht, The Netherlands
[2]Department of Neurology, Vivantes Klinikum im Friedrichshain, Berlin, Germany

We meta-analytically reviewed studies that used the Structured Inventory of Malingered Symptomatology (SIMS) to detect feigned psychopathology. We present weighted mean diagnostic accuracy and predictive power indices in various populations, based on 31 studies, including 61 subsamples and 4009 SIMS protocols. In addition, we provide normative data of patients, claimants, defendants, nonclinical adults, and various experimental feigners, based on 41 studies, including 125 subsamples and 4810 SIMS protocols. We conclude that the SIMS (1) is able to differentiate well between instructed feigners and honest responders; (2) generates heightened scores in groups that are known to have a raised prevalence of feigning (e.g., offenders who claim crime-related amnesia); (3) may overestimate feigning in patients who suffer from schizophrenia, intellectual disability, or psychogenic non-epileptic seizures; and (4) is fairly robust against coaching. The diagnostic power of the traditional cut scores of the SIMS (i.e., > 14 and > 16) is not so much limited by their sensitivity—which is satisfactory—but rather by their substandard specificity. This, however, can be worked around by combining the SIMS with other symptom validity measures and by raising the cut score, although the latter solution sacrifices sensitivity for specificity.

INTRODUCTION

In 1997 Smith and Burger introduced their *Structured Inventory of Malingered Symptomatology* (SIMS). Its name is misleading, but clearly reveals the ambition of the authors: The SIMS aims to detect feigned psychopathology. Its construction was guided by the idea that naïve respondents are likely to endorse bizarre, rare, atypical, or extreme symptoms on a questionnaire when they attempt to feign or exaggerate symptoms. Thus, the SIMS presents patients, claimants, defendants, or research participants with a list of 75 implausible symptoms or statements that are to be endorsed or rejected.

The SIMS covers a broad spectrum of pseudopsychopathology. Its items allude to atypical depression, improbable memory problems, pseudoneurological symptoms, doubtful claims of psychotic experiences, and hyperbolic signs of mental retardation. Each of these five categories is represented by a subscale comprising 15 items. People

may endorse some items, but those who claim to suffer from an abundance of SIMS items are thought to feign psychopathology. Smith and Burger (1997) recommended a cut score of > 14 (i.e., scores *exceeding* 14 are considered to be positive test outcomes). The authors warned that the SIMS subscales are not suitable for detecting feigned psychopathology, and only serve to evaluate what type of psychopathology the respondent is trying to feign, once it has been established that the total SIMS score exceeds the cutoff.

In 2005, after the SIMS had been in use for a number of years, its manual was published (Widows & Smith, 2005). Studies that had employed the SIMS to that date were summarized in the manual. Boone (2013), Wisdom, Callahan, and Shaw (2010; Table 1), and Smith (2008; Table 19.3) presented qualitative reviews of studies using the SIMS. In this article we take a meta-analytical approach to the extant SIMS literature. In contrast to previous reviews, we present a meta-analysis of diagnostic accuracies and predictive powers, and offer normative data of patients, claimants, defendants, nonclinical adults, and various experimental feigners. We focus on what these data tell us about the ability of the SIMS to discriminate between feigners and honest responders. We also discuss implications for the clinical utility of the SIMS and offer guidelines for clinical practice. Before doing so, we put the term "malingering test" (as the SIMS is often understood) into context.

Malingerers over-report pathological symptoms. They do so in a calculated attempt to obtain material gain or to escape formal duty or responsibility (see *Diagnostic and Statistical Manual of Mental Disorders,* 5th ed.; *DSM–5*; American Psychiatric Association, 2013). However, not everyone who over-reports symptoms is a malingerer. Hence, there is no such thing as a malingering test. While a test may indicate that a person over-reports symptoms, it cannot clarify *why* he or she does so (see also Boone, 2007). Thus, the name of the SIMS is based on a misconception. It is more accurate to refer to the SIMS as a symptom validity test (SVT).

How proficient is the SIMS in detecting symptom over-endorsement? Several authors (e.g., Hartman, 2002) have commented on the qualities that SVTs such as the SIMS should possess. First and foremost, an SVT should differentiate accurately between honest respondents and people who are known to feign their symptoms. This requires high sensitivity (i.e., ability to identify feigners) as well as high specificity (i.e., ability to classify honest responders correctly). A second and related point is that an SVT should be sensitive to differential prevalence. That is, the SVT should generate higher scores for populations in which feigning is common, and lower scores for populations in which feigning is rare. A third requisite for SVTs is insensitivity to genuine psychopathology: Honest patients should not attain red flag scores on an SVT. A fourth quality indicator is robustness against coaching by, for example, lawyers. Other criteria stress that SVTs should be easy to administer and interpret, and that they ought to cover disabilities or symptoms that are likely targets for feigning. Ideally, an SVT resembles measures of genuine psychopathology (i.e., it should not be readily identifiable as an SVT). Lastly, the efficacy of an SVT should be based on reliable, up-to-date norms for nonclinical controls, honest patients, and known feigners (Hartman, 2002).

Below we review the SIMS literature and match the SIMS against these criteria. We conclude that while the SIMS might be a useful SVT, certain guidelines should be adopted. These include using optimal cut scores, exercising extra caution when administering the SIMS to certain clinical groups, and combining the SIMS with other SVTs

and, ideally, with performance validity tests (PVTs), which measure underperformance in cognitive domains.

METHOD

Published studies that employed the SIMS were located by means of a computerized literature search with Google Scholar and EBSCO Discovery Service (PsycINFO). This search was conducted with "Structured Inventory of Malingered Symptomatology" and "SIMS" as entries. Studies published since the introduction of the SIMS (1997) up to the writing of this article (August 2014) were inspected. In addition to Anglo-Saxon articles, Dutch, German, and Spanish papers were included. Unpublished papers (e.g., dissertations) were not considered because they are not readily accessible to readers. Accordingly, five dissertations were not included. Additional studies were identified by contacting researchers who attended the Third European Symposium on Symptom Validity Assessment (see Plohmann & Merten, 2013).

All identified studies that reported data concerning the diagnostic accuracy and/or predictive power of the SIMS in detecting symptom over-reporting were included (i.e., no exclusions were made). These studies are listed in Table 1 (known-groups studies) and Table 2 (simulation studies). Weighted mean diagnostic accuracy and predictive power indices for both known-groups and simulation studies are presented in Table 3. Table 6 gives mean SIMS scores and effect sizes, based on studies that reported mean SIMS scores. In order to test whether language had a moderating effect, all studies were coded with respect to their language background. The effect sizes shown in Table 6 were calculated by employing a single pooled standard deviation. Data were analyzed with IBM SPSS Statistics 21 and Exploratory Software for Confidence Intervals (ESCI; see Cumming, 2012).

RESULTS

Known-groups studies

Table 1 summarizes SIMS data from 10 studies in which an external criterion was used to define samples of feigners and honest responders. With the exception of González Ordi, Santamaría Fernández, and Fernández Marín (2010), samples consisted of individuals who were involved in legal procedures, be it as claimants in compensation cases, as defendants in criminal procedures, or as prisoners within a penal institution. In most studies, scores on the *Structured Interview of Reported Symptoms* (SIRS; Rogers, Bagby, & Dickens, 1992) served as external criteria. The SIRS consists of 172 questions that cover multiple strategies to detect feigned psychopathology, such as absurd symptoms, unlikely combinations of symptoms, reported versus observed symptoms, and abnormal severity of symptoms. The SIRS has been well studied; the meta-analysis by Green and Rosenfeld (2011) yielded a sensitivity (i.e., the likelihood of a positive SVT result in feigners) of .49 and a specificity (i.e., likelihood of a negative SVT result in honest responders) of .95 for the SIRS.

The overall conclusion that can be drawn from Table 1 is that the SIMS is fairly effective in discriminating between feigning and honest responding groups: Effect sizes

Table 1. Diagnostic accuracy of the Structured Inventory of Malingered Symptomatology (SIMS) in known-groups studies

Authors	Sample	Sample size	Criterion	Prevalence of feigning	Cutoff	Sensitivity	Specificity	Cohen's d
Alwes et al. (2008)	Claimants	308	SIRS / TOMM, VSVT, LMT	Psychiatric: 7% / Cognitive: 24% (Total: 31%)	>16 / >16	.96 / .75	.67 / .60	2.6 / 1.1
Clegg et al. (2009)	Claimants	56	SIRS	31%	>14 / >16 / >19	1.00 / .95 / .90	.37 / .51 / .69	1.6
Edens et al. (2007)	Prisoners in mental health unit	56	SIRS	45%	>14	1.00	.52	NR
			Clinical judgment based on SIRS strategies	46%	>14	.85	.40	NR
González Ordi, Santamaría Fernández, Blasco Saiz, & Pallardó Durá (2008)	Claimants	305	MMPI-2 (F-K ≥ 5 or T and F > 70)	10%	>14 / >16	1.00 / .97	.61 / .75	2.1
González Ordi et al. (2010)	Patients on sick leave	61	Independent judgment similar to Slick	56%	>14 / >16	.88 / .82	.81 / .93	1.6
Heinze & Purisch (2001)	Defendants incompetent to stand trial	57	Clinical judgment by multi-disciplinary team	13%[a]	>13	.87	N/A	N/A
Lewis et al. (2002)	Defendants	55	SIRS	44%	>16	1.00	.61	3.0
Vitacco et al. (2007)	Defendants	100	SIRS	21%	>14	1.00	.65	3.1
Vossler-Thies et al. (2013)	Claimants	95	Slick	Definite: 31% / Probable: 22% / Total: 53%	>16 / >16 / >16	NR / NR / NR	NR / NR / NR	1.5[b] / 1.1[b] / 1.3[b]
Wisdom et al. (2010)	Claimants	33	Slick	67% (15% definite, 52% probable)	>14 / >21 / >24	.96 / .68 / .55	.64 / .73 / 1.00	2.0[b] / 1.5[b] / 1.5[b]

NR = Not reported, insufficient data to calculate. N/A = Not applicable. LMT = Letter Memory Test (Inman et al., 1998). MMPI-2 = Minnesota Multiphasic Personality Inventory–Second Edition (Butcher et al., 2001). SIRS = Structured Interview of Reported Symptoms (Rogers et al., 1992). Slick = criteria for malingering as formulated by Slick, Sherman, and Iverson (1999). TOMM = Test of Memory Malingering (Tombaugh, 1996). VSVT = Victoria Symptom Validity Test (Slick, Hopp, Strauss, & Spellacy; 1996). [a]Based on an initial sample of 438 defendants who were found incompetent to stand trial. [b]Calculated using available information.

Table 2. Diagnostic accuracy of the Structured Inventory of Malingered Symptomatology (SIMS) in simulation studies

Authors	Sample	Experimental feigners	Honest responders	Cutoff	Sensitivity	Specificity	Cohen's d
Cima et al. (2003)	Patients	–	62	>16	N/A	.73	1.5[d]
	Undergraduates	135	69	>16	.87	1.00	2.0
Dandachi-FitzGerald & Merckelbach (2013)	Undergraduates	42	20	>16	.91	1.00	3.2[d]
Edens et al. (1999)	Undergraduates	196	196[a]	>14	.96	.91	3.0[d]
Edens et al. (2007)	Nonclinical prison inmates	29	30	>14	.90	.97	NR
Giger et al. (2010)	Nonclinical adults:						
	– Naïve	20	20	>16	.95	.95	2.9[d]
	– Forewarned of SVT	20	–	>16	.65	N/A	1.5[d]
González Ordi & Santamaría Fernández (2008)	Patients	–	152	>14	.92	.40	1.4
	Undergraduates	326[c]	–	>16	.96	.48	
González Ordi, Santamaría Fernández, & Matalobos Veiga (2008)	Undergraduates	326[c]	147	>14	.96	.95	3.0
				>16	.92	.97	
Graue et al. (2007)	Intellectually disabled	–	26	>16	N/A	.23	.5
	Matched controls	25	10	>16	.88	NR	1.3[d]
Jelicic et al. (2006)	Undergraduates:						
	– Naïve	15[b]	15	>16	.93	1.00	3.2[d]
	– Clinical knowledge	15[b]	–	>16	1.00	N/A	3.0[d]
	– Knowledge + forewarned	15[b]	–	>16	.80	N/A	2.6[d]
	(Total)	(45)	(15)	>16	(.91)	(1.00)	(2.9)[d]
Jelicic et al. (2011)	Undergraduates:						
	– Clinical knowledge	29[b]	29	>16	.93	1.00	3.0[d]
	– Knowledge + forewarned	28[b]	–	>16	.86	N/A	3.2[d]
	(Total)	(47)	(29)	>16	(.89)	(1.00)	(3.1)[d]
Jelicic, Merckelbach, et al. (2007)	Undergraduates:						
	– Naïve	30	30	>14	.90	1.00	2.8[d]
	– Knowledge + forewarned	30	–	>14	.90	N/A	2.3[d]
Jelicic, Peters, et al. (2007)	Undergraduates:						
	– Naïve	30	30[a]	>16	.97	1.00	4.1[d]
	– Clinical knowledge	31	31[a]	>16	1.00	1.00	3.4[d]
Jelicic et al. (2013)	Clinical experts	23	23[a]	>16	.87	1.00	2.4[d]
	Undergraduates	24	24[a]	>16	.96	1.00	3.7[d]
Merckelbach & Collaris (2012)	Undergraduates:						
	– Naïve	15[b]	15	>16	NR	NR	2.1[d]

(*continued*)

Table 2. (Continued)

Authors	Sample	Experimental feigners	Honest responders	Cutoff	Sensitivity	Specificity	Cohen's d
	– Forewarned of SVT	15[b]	15	>16	NR	NR	2.3[d]
	– Primed for honesty	15[b]	15	>16	NR	NR	2.3[d]
	(Total)	(45)	(45)	>16	(.76)	(1.00)	(2.2)[d]
Merckelbach & Smith (2003)	Patients	–	10	>16	N/A	NR	2.1[d]
	Undergraduates	57	231	>16	.93	NR	4.7[d]
	(Total)	(57)	(241)	>16	(.93)	(.98)	(4.7)
Merten et al. (2008)	Nonclinical adults:						
	– Naïve	16	–	>16	.94	N/A	N/A
	– Forewarned of SVT	16	–	>16	.63	N/A	N/A
Merten et al. (2010)	Nonclinical adults:						
	– Naïve	20[b]	–	>16	.90	N/A	N/A
	– Clinical knowledge	20[b]	–	>16	.90	N/A	N/A
	– Forewarned of SVT	20[b]	–	>16	.85	N/A	N/A
	– Knowledge + forewarned	20[b]	–	>16	.70	N/A	N/A
	(Total)	(80)	–	>16	(.84)	N/A	N/A
Peters et al. (2013)	Schizophrenia patients	–	41	>16	N/A	.71	1.3[d]
	Nonclinical controls	–	43	>16	N/A	1.00	
Rogers et al. (1996)	Forensic patients:	–	53[a]	>16	N/A	.83[d]	NR
	– Clinical knowledge	53	–	>16	.70[d]	N/A	
Rogers et al. (2005)	Undergraduates:						
	– Malingering	17[b]	16	>16	NR	NR	1.3[d]
	– Factitious, Demanding	18[b]	–	>16	NR	NR	1.1[d]
	– Factitious, Dependent	14[b]	–	>16	NR	NR	0.9[d]
	(Total)	(49)	–	>16	NR	NR	(1.1)[d]
Rogers et al. (2014)	Inpatient trauma victims	53	54	>14	.93	.28	1.7
Smith & Burger (1997)	Undergraduates	208	30	>14	.96	.88	2.1
Vossler-Thies et al. (2013)	Nonclinical adults	25	–	>16	.92	N/A	N/A
Zimmermann et al. (2013)	Military patients	–	96	>16	N/A	.72	3.4[d]
	Military personnel	40	–	>16	1.00	N/A	

NR = Not reported, insufficient data to calculate. N/A = Not applicable. [a]Within-participants (i.e., test–retest) design. [b]The difference in mean SIMS scores between these groups was not statistically significant ($p > .05$). [c]The data of this group were employed by both González Ordi and Santamaría Fernández (2008) and González Ordi, Santamaría Fernández, and Matalobos Veiga (2008). [d]Calculated using available information.

Table 3. Weighted mean diagnostic accuracy and predictive power of the Structured Inventory of Malingered Symptomatology (SIMS) in various samples, based on 31 studies, including 61 subsamples and 4009 SIMS protocols

Population	Number of samples	Total sample size	Cutoff	Sensitivity	Specificity	Positive predictive power at base rate:			Negative predictive power at base rate:		
						10%	30%	50%	10%	30%	50%
Claimants	4	455	>14	.98	.61	.22	.52	.72	1.00	.99	.97
(Known-groups)	4	730	>16	.91	.70	.25	.57	.75	.99	.95	.89
Defendants	2	156	>14	1.00	.60	.22	.52	.71	1.00	1.00	1.00
(Known-groups)	2	108	>16	1.00[b]	.67	.25	.56	.75	1.00	1.00	1.00
Honest patients	2	206	>14	N/A	.37	.15[e]	.40[e]	.61[e]	.99[e]	.97[e]	.93[e]
	5[a]	377	>16	N/A	.59	.20[f]	.49[f]	.69[f]	.99[f]	.95[f]	.88[f]
Nonclinical adults	10	1222[c]	>14	.96	.93	.60	.85	.93	1.00	.98	.96
(Experimental groups)	25	1487[d]	>16	.92	.99	.91	.98	.99	.99	.97	.93
Experimental feigners: – Clinical knowledge	6	171	>16	.87	N/A	.22[g]	.52[g]	.72[g]	.98[g]	.92[g]	.84[g]
Experimental feigners: – Forewarned of SVT	3	56	>16	.72	N/A	.19[g]	.48[g]	.68[g]	.95[g]	.85[g]	.70[g]
Experimental feigners: – Knowledge + forewarned	4	88	>16	.82	N/A	.21[g]	.51[g]	.71[g]	.97[g]	.90[g]	.79[g]

The data summarized in this table are based on 3652 individuals who—due to within-participant (i.e., test–retest) designs—contributed 4009 SIMS protocols. The studies used are listed in Tables 1 and 2. The studies by Heinze and Purisch (2001) and Rogers, Jackson, and Kaminski (2005) could not be included in this table: The former because it reported only data for a cutoff of >13; the latter for the reason that it did not report diagnostic accuracy data. [a]The honesty of these patients is assumed, but not established through symptom validity testing. [b]Based on 1 sample, 55 participants. [c]This sample includes 789 SIMS protocols of experimental feigners and 433 SIMS protocols of nonclinical controls. [d]This sample includes 750 SIMS protocols of experimental feigners and 737 SIMS protocols of nonclinical controls. [e]Calculated by assuming a sensitivity for feigning responders of .97 (the weighted mean sensitivity of cut score >14 for experimental feigners and feigning claimants and defendants). [f]Calculated by assuming a sensitivity for feigning responders of .92 (the weighted mean sensitivity of cut score >16 for experimental feigners and feigning claimants and defendants). [g]Calculated by assuming a specificity for honest responders of .66 (the weighted mean specificity of cut score >16 for honest patients, claimants, and defendants).

(Cohen's *d*s) range from 1.1 to 3.0. The sensitivity for the commonly employed cutoffs (i.e., > 14 and > 16) is adequate, ranging from .75 to 1.00. The corresponding specificity rates are highly divergent (range: .37 to .93), yet often alarmingly low; a serious point to which we will return below.

Simulation studies

Table 2 provides an overview of 24 simulation studies. Samples consisted predominantly of undergraduate students. In the majority of these studies participants were either asked to feign psychopathology or to respond honestly. In the feigning conditions participants were typically presented with a case vignette describing an individual with a salient and strong motive to feign pathological symptoms. Before being administered the SIMS, participants were prompted to identify with the main character in the case scenario and were asked to feign psychiatric or cognitive symptoms convincingly.

In most studies participants feigned symptoms on the basis of their own naïve views. However, some studies employed conditions in which participants were informed about genuine and noncredible symptoms. Still other studies recruited participants who already possessed such knowledge by virtue of their professional or psychiatric background. A number of studies explicitly added a warning about validity tests and information about the rationale behind tests such as the SIMS. Providing participants with such technical information amounts to *coaching*, an issue that we will consider below.

The general conclusion that can be drawn from Table 2 is, again, that the SIMS does a reasonably good job in differentiating between experimental feigners and honest responders: Effect sizes (Cohen's *d*s) vary between .5 and 4.7. The range of effect sizes is remarkably wide. This is in large part due to differences in type of control groups. Some studies employed patients as controls (e.g., González Ordi & Santamaría Fernández, 2008; Graue et al., 2007; Peters, Jelicic, Moritz, Hausschildt, & Jelinek, 2013), whereas others resorted to nonclinical controls. This variety in symptomatological levels will have contributed to differences in SIMS scores between control groups. Additionally, some conditions may render patients particularly susceptible to produce raised SIMS scores (e.g., intellectual disability; Graue et al., 2007; see below).

As can be seen in Table 6 later, nonclinical controls tend to produce significantly lower SIMS scores than patient controls, meaning that the SIMS is sensitive to genuine psychopathology. This inflates effect sizes and specificity values (see also Table 3). In nonclinical control groups specificity rates of commonly used cutoffs (i.e., > 14 and > 16) vary from .88 to 1.00, whereas in patient controls they range from .23 to .83. Compared with specificity rates, sensitivity rates are more satisfactory and less diverse across studies. Table 2 lists sensitivity rates for simulation studies, which ranged from .87 to 1.00 for naïve (i.e., *noncoached*) groups.

Robustness against coaching

Several studies tested whether the sensitivity of the SIMS to detect feigning is undermined by knowledge about psychopathology and/or advise to beware of symptom validity testing. These studies are also listed in Table 2, and their accumulated data are displayed in Table 3.

The general finding is that being knowledgeable about psychopathology does little to undermine the sensitivity of the SIMS. In contrast, forewarning of symptom validity testing or advice against excessive feigning undermines sensitivity: When individuals were forewarned about SVTs, a cut score of > 16 yielded a sensitivity of only .72. Furthermore, clinical knowledge interacts significantly with forewarning, but in a counterintuitive way. That is, knowledge about psychopathology reduces the sensitivity-undermining effect of forewarning by approximately one third.

The SIMS attempts to detect over-reporting of psychopathology, which is radically different from the type of symptom validity assessed by PVTs. PVTs such as the Amsterdam Short-Term Memory test (ASTM; Schagen, Schmand, de Sterke, & Lindeboom, 1997), the Test of Memory Malingering (TOMM; Tombaugh, 1996), and the Word Memory Test (WMT; Green, 2003) focus on cognitive dysfunction and measure underperformance on cognitive tests. Nonetheless, SIMS scores are significantly elevated in groups that are instructed to feign cognitive deficits. For example, the SIMS

attained a higher sensitivity than the ASTM among respondents who were coached and asked to feign cognitive impairment (detection rates of .90 and .70, respectively; Jelicic, Merckelbach, Candel, & Geraerts, 2007). The sensitivity of the SIMS matched that of the ASTM in a study on the effects of forewarning among experimental feigners of cognitive dysfunction (Giger, Merten, Merckelbach, & Oswald, 2010). In addition, Jelicic, Ceunen, Peters, and Merckelbach (2011) found the SIMS to be as sensitive as the TOMM when coached feigners simulated cognitive problems (detection rates of .87 and .86, respectively). In a simulation study by Merten, Diederich, and Stevens (2008), the SIMS was more sensitive in detecting feigned whiplash injury symptoms than the WMT. Yet, in forewarned participants it showed a drop of .31 in detection rate, while the WMT was largely resistant against forewarning. Moreover, 69% of research participants were able to identify the SIMS as an instrument that assesses symptom validity, as opposed to 56% who were able to identify the WMT as such (Merten et al., 2008). In another study, however, only 16% of the participants suspected the SIMS to be a measure of symptom validity, while 28% voiced such suspicion about the WMT (Merten, Lorenz, & Schlatow, 2010).

These divergent findings illustrate that estimates of sensitivity, transparency, and robustness against coaching depend on the experimental procedures that participants are subjected to. The sensitivity of the SIMS to feigned cognitive dysfunction might be largely due to the tendency of some feigners to overgeneralize and simulate psychopathology in addition to cognitive deficits. However, except for low intelligence, the SIMS does not address specific feigned cognitive deficits. Thus, regardless of its potential sensitivity, the SIMS cannot be relied upon to detect feigned cognitive impairment.

Diagnostic accuracy

Table 3 shows weighted mean diagnostic accuracy and predictive power of the SIMS for known-groups research (Table 1) and simulation studies (Table 2). Table 3 allows for the following conclusions. To begin with, using a cut score of > 14 (as recommended by Smith & Burger, 1997) results in a sensitivity that circles around .97. Even when the cutoff is raised to > 16, the sensitivity of the SIMS to detect feigned symptoms, such as displayed by defendants attempting to feign insanity or incompetency to stand trial, remains excellent (i.e., above .90). The sensitivity rates in experimental feigners (Table 2) are similar to those in claimants (Table 1). This suggests that experimental feigners are a valid model for claimants and defendants who feign symptoms; a point that is further elaborated upon below.

Second, specificity rates in samples of honest patients, claimants, and defendants using cut scores of > 14 and > 16 varied extensively (.37 to .70) and indicate that the SIMS can yield high false positive rates in these groups (i.e., misclassify honest responders as feigners). Specificity rates in groups of nonclinical controls are high, but not perfect, which is disappointing given that these respondents presumably do neither feign nor experience any form of psychopathology. The marked difference in specificity rates between clinical and nonclinical controls underscores the importance of employing samples of honest patients, claimants, or defendants as control groups when estimating diagnostic accuracy indices of SVTs.

Third, although some studies found that control patients with psychotic symptoms score relatively low on the SIMS (e.g., Vitacco, Rogers, Gabel, & Munizza,

2007), other studies noted that the diagnostic accuracy of the SIMS appears to be limited in patients with schizophrenia (Peters et al., 2013) and individuals who suffer from psychogenic non-epileptic seizures (Benge et al., 2012) or intellectual disability (Graue et al., 2007). The nontrivial proportion of patients with scores above the cutoff suggests that the lack of specificity is overly pronounced in these clinical groups.

Estimates of the diagnostic accuracy of the SIMS are dependent on the criteria that are employed to compose groups of feigners and honest responders. The majority of known-groups studies employed the SIRS as sole external criterion, but this instrument possesses suboptimal sensitivity. Because of this, groups of honest responders may contain a substantial proportion of false negatives (i.e., feigners who are erroneously classified as honest responders; on average 51%; Green & Rosenfeld, 2011). This may lead to an overestimation of the sensitivity of the SIMS (because the SIMS is not faulted for failing to label false negatives of the SIRS as hits) and an underestimation of specificity (as the SIMS is considered to be incorrect when it does identify false negatives of the SIRS as hits).

A related problem with using the SIRS as a criterion is the similarity in detection strategies between the SIMS and the SIRS. Both instruments focus on the endorsement of bizarre, extreme, or atypical symptoms, which is reflected in the significant correlations between the SIMS and the SIRS (e.g., $r = .54$, Freeman, Powell, & Kimbrell, 2008; and $r = .81$, Edens, Poythress, & Watkins-Clay, 2007). Thus, the diagnostic errors of the SIMS may mirror those of the SIRS, which might lead to an overestimation of the diagnostic accuracy of the SIMS.

Predictive power

Diagnostic accuracy (i.e., sensitivity and specificity) is only one part of the equation that describes the efficacy of a diagnostic instrument. The other part is predictive power, which depends on both diagnostic accuracy and the prior probability (i.e., prevalence, or base rate) of test results. Where diagnostic accuracy denotes the probability of a certain test outcome given the true status of an individual, predictive power refers to the probability of the true status of an individual given a certain test outcome. Thus, the practical implications of particular SIMS cutoffs cannot be evaluated without taking the base rate of feigning into account.

Table 3 shows predictive power for base rates of 10%, 30%, and 50%. Higher base rates lead to greater positive predictive power (i.e., probability of feigning if SVT outcome is positive) for a given cutoff, whereas lower base rates lead to a greater negative predictive power (i.e., probability of honest responding if SVT outcome is negative). The negative predictive power for cut scores of > 14 and > 16 is excellent, even at base rates of up to 50%. That is, the likelihood that an individual *does not* feign symptoms if his or her SIMS score remains below the cutoff is very high. What is more, negative predictive power decreases only slightly as base rate increases, especially if a cutoff of > 14 is employed. On the other hand, the likelihood that an individual feigns symptoms if his or her SIMS score exceeds the cutoff varies considerably depending on the base rate of feigning in his or her population. Overall, the positive predictive power for cut scores of > 14 and also > 16 is rather low, and drops

quickly as base rates fall. This implies that in populations where feigning is rare, the probability of false-positive identifications is high, even when a cut score of > 16 is employed. Note that positive predictive power is markedly higher among nonclinical respondents than in patient populations. As said earlier, the SIMS is sensitive to genuine psychopathology, and can overestimate feigning in patients.

Sensitivity to differential prevalence

Table 4 provides an overview of five studies that employed a differential design to explore the qualities of the SIMS. The samples vary from neurological patients and fibromyalgia litigants to prisoners who claim crime-related amnesia. The idea that feigned psychopathology is more prevalent among patients who report psychogenic complaints than among patients with neurological symptoms is *prima facie* plausible. The same holds for the assumption that feigning is more common among defendants who claim crime-related amnesia than it is among defendants without such claims. Similarly, it is safe to assume that feigning occurs more frequently among litigants than among nonlitigants.

Table 4 illustrates that the mean SIMS score of litigants is substantially higher than that of nonlitigating patients (Capilla Ramírez, González Ordi, & Santamaría Fernández, 2008). Furthermore, SIMS scores above the cutoff are more prevalent in samples of patients with psychogenic complaints or crime-related amnesia by an average factor of 4.9 in comparison to relevant control groups. Taken together, the data suggest that the SIMS is sensitive to differential prevalence. Moreover, this sensitivity seems to manifest itself more strongly as the cutoff is raised (Benge et al., 2012).

Correlations with clinical scales and other SVTs

Table 5 lists Pearson product–moment correlations between SIMS scores and scores on other psychological instruments. It warrants the following conclusions. First, respondents who endorse many symptoms on the SIMS also report many symptoms on standard clinical inventories: There are moderate correlations (rs = .50 to .72) between the SIMS and symptom measures such as the Symptom Checklist-90 Revised (SCL-90-R; Derogatis, 1994) and the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961).

Second, the correlations between the SIMS and PVTs (ASTM; Schagen et al., 1997; Morel Emotional Numbing Test, MENT; Morel, 1998; WMT; Green, 2003) are weak to moderate (rs vary from .22 to .49). Assessment of symptom validity should ideally include multiple SVTs, and preferably SVTs that are independent (i.e., correlate weakly with each other; Larrabee, 2008, 2014; Victor, Boone, Serpa, Buehler, & Ziegler, 2009). With this in mind, the correlations listed in Table 5 suggest that the combination of SIMS and ASTM is less redundant compared with that of SIMS and WMT.

A final point is the interpretation of SIMS scores that exceed the cutoff. In the studies listed in Tables 4 and 5, two extremes can be found. Some authors employ the SIMS as a definitive malingering test and treat scores above the cutoff as conclusive evidence of feigned psychopathology (e.g., Kunst, Winkel, & Bogaerts, 2011). Others

consider heightened SIMS scores to be a manifestation of somatization (e.g., Benge et al., 2012). Both interpretations are questionable. The first position is dubious because heightened SIMS scores do not necessarily reflect malingering. Elevated SIMS scores may also signal feigned psychopathology as found in factitious disorder or they might be the result of yea-saying to the test items. Such irrelevant responding can stem from frustration, boredom, defiance, or fatigue, without flagging feigning *per se* (see Meade & Craig, 2012).

The second interpretation assumes that somatoform disorders are superordinate in relation to feigning. There is no empirical evidence for this assumption, and the opposite idea—intentional feigning eventually contributing to somatoform symptoms—can be defended equally forcefully (Merten & Merckelbach, 2013; see also Rogers, Jackson, & Kaminski, 2005).

SIMS subscales

Feigning may take on many forms. One advantage of the SIMS is that it comprises five non-overlapping subscales that cover diverse types of pseudopsychology: Psychosis, Neurological Impairment, Amnestic Disorders, Low Intelligence, and Affective Disorders. However, the diagnostic accuracy indices of these subscales are inferior to that of the SIMS total score. Among various experimental samples, the original cutoffs of the subscales produced sensitivity rates between .66 and 1.00 and

Table 4. Differential prevalence of Structured Inventory of Malingered Symptomatology (SIMS) scores in clinical and forensic samples

Authors	Samples	Sample size	Cutoff	Prevalence of SIMS scores beyond cutoff	Mean SIMS scores	Cohen's d
Benge et al. (2012)	Patients with psychogenic non-epileptic seizures	91	>16 (>19)	71% (59%)	22.4[c]	.8[d]
	Patients with epileptic seizures	29	>16 (>19)	31% (17%)	14.4[c]	
Cima & van Oorsouw (2013)	Prisoners claiming crime-related amnesia	12	>16	33%	12.5[b]	.8
	Non-amnestic prisoners	19	>16	5%	6.9[b]	
Chen et al. (2011)	Successful versus unsuccessful placebo-induced seizures in patients with psychogenic non-epileptic seizures	51	>14	Successful: 80%[a] Unsuccessful: 33%[a]	NR	NR
Capilla Ramírez et al. (2008)	Fibromyalgia claimants	30	>16	80%	25.5[c]	2.5[d]
	Chronic pain claimants	30	>16	NR	14.1[c]	1.0[d]
	Chronic pain nonclaimants	25	>16	NR	6.9[c]	
van Beilen et al. (2009)	Psychogenic patients	26	>16	23%	11.5[c]	1.3[d]
	Neurological patients	26	>16	4%	7.8[c]	.6[d]
	Nonclinical controls	18	>16	0%	4.6[c]	

NR = Not reported, insufficient data to calculate. [a]Difference is significant at a level of $p < 0.05$ (two-tailed). [b]Difference is significant at a level of $p < 0.01$ (two-tailed). [c]Difference is significant at a level of $p < 0.001$ (two-tailed). [d]Calculated using available information.

Table 5. Pearson product–moment correlation coefficients between Structured Inventory of Malingered Symptomatology (SIMS) scores and scores of other instruments

Authors	Samples	Sample size	Prevalence of feigning	Measure	Correlation with SIMS	Shared variance
Cima & van Oorsouw (2013)	Prisoners claiming crime-related amnesia	12	33%	PPI	.44	19%
	Non-amnestic prisoners	19	5%			
Dandachi-FitzGerald et al. (2011)	Psychiatric outpatients	183	21%	ASTM	−.22[a]	5%
Edens et al. (2007)	Prisoners in mental health unit	56	45%	SIRS	.81	66%
Freeman et al. (2008)	Veterans with PTSD	74	53%	SIRS	.54	29%
Kunst et al. (2011)	Claimants victim services	125	18%	PTSD	.67	45%
Merten, Friedel, Mehren, & Stevens (2007)	Claimants workers' compensation	93	26%	BDI	.72	52%
				WMT	−.44[a]	19%
Merten, Thies, Schneider, & Stevens (2009)	Claimants workers' compensation	61	51%	MENT	.36	13%
				WMT	−.49[a]	24%
van Beilen et al. (2009)	Psychogenic patients	26	23%	SCL-90-R	.70	49%
	Neurological patients	26	4%		.50	25%
	Nonclinical controls	18	0%		Not significant	

ASTM = Amsterdam Short-Term Memory test (Schagen et al., 1997). BDI = Beck Depression Inventory (Beck et al., 1961). MENT = Morel Emotional Numbing Test (Morel, 1998). PPI = Psychopathic Personality Inventory (Lilienfeld & Andrews, 1996). PTSD = Post-Traumatic Stress Disorder symptoms. SCL-90-R = Symptom Checklist-90-Revised (Derogatis, 1994). SIRS = Structured Interview of Reported Symptoms (Rogers et al., 1992). WMT = Word Memory Test (Green, 2003). [a]Negative correlation is due to the inverse scoring direction of the two instruments.

specificity rates between .73 and .93, with two notable exceptions: The sensitivity of the Psychosis subscale, the lowest of which was .57 and the specificity of the Low Intelligence subscale, the lowest of which was .52 (Edens, Otto, & Dwyer, 1999; Smith & Burger, 1997). However, these diagnostic accuracy rates are based on nonclinical controls and therefore overestimate specificity values as found in clinical samples.

The sensitivity of the subscales seems to be largely dependent on the type of symptoms that are feigned. The closer the match between the type of psychopathology a respondent tries to feign and the type of pseudopsychopathology that a subscale targets, the higher the sensitivity of the subscale (Merckelbach & Smith, 2003; Smith & Burger, 1997; but see Edens et al., 1999, Table 3). While the sensitivity of a SIMS subscale might thus benefit from certain contexts, it rarely exceeds that of the SIMS total score. Subscales that attain high sensitivity when confronted with their target psychopathology are Affective Disorders, Neurological Impairment, and Amnestic Disorders (Benge et al., 2012; Clegg, Fremouw, & Mogge, 2009; Giger et al., 2010). The Low Intelligence subscale remains relatively insensitive to explicit attempts to mimic poor intellectual abilities (Clegg et al., 2009).

Cronbach's alpha coefficients (taken to be measures of internal consistency, but see Sijtsma, 2009) of the subscales depend on sample characteristics. Smith and Burger (1997) obtained alpha coefficients ranging from .80 to .86 in a sample in which the majority (87%) was instructed to *feign symptoms* (N = 476 undergraduates), whereas Merckelbach and Smith (2003) found coefficients that varied between .24 (Low Intelligence) and .59 (Affective Disorders) in a sample where the majority (81%) was asked to *respond honestly* (N = 298 undergraduates). Alpha values are higher in feigning than in honest conditions; a pattern that is also evident in the data reported by Rogers, Robinson, and Gillard (2014). The influence of sample characteristics is further illustrated by the discrepancies between the results of Merckelbach and Smith (2003) and Vitacco et al. (2007; N = 100 competency to stand trial patients, 21% malingering). The alpha coefficients for the Affective Disorders subscale in these studies were .59 and .61, respectively. Yet in the former study it was the highest coefficient for subscales, while in the latter it was the lowest. Poor alphas for the Affective Disorders subscale (.31–.37) were also found by Rogers et al. (2014) in their clinical samples. Studies uniformly find that the alpha coefficients of the subscales are lower than that of the total scale, which is to be expected given that internal reliability is a function of item number. The internal consistency of the SIMS total scale is generally satisfactory (e.g., α = .80 in Cima et al., 2003; .72 in Merckelbach & Smith, 2003; .92–.94 in Rogers et al., 2014; and .96 in Vitacco et al., 2007).

All in all, the subscales do not lend themselves to detect feigned psychopathology. Still, the subscales are sensitive enough to their target psychopathology to justify qualitative use (i.e., tentative assessment of the type of psychopathology that a person is trying to feign, once it has been established that the total score exceeds the cutoff), as originally proposed by Smith and Burger (1997).

The SIMS covers a wide variety of (spurious) psychopathology. Nonetheless, it exhibits sufficient internal consistency (e.g., Vitacco et al., 2007). This might reflect a common strategy among feigners—the tendency to overgeneralize when reporting symptoms. Studies involving experimental feigners typically find that, in spite of instructions to feign specific symptoms, participants are inclined to overgeneralize and report a diversity of symptoms (e.g., Giger et al., 2010; Merten et al., 2010).

Clegg et al. (2009) noted that experimental feigners who were instructed to feign depression and patients who were suspected of feigning a mood disorder were just as likely to fail the SIMS as they were to fail the Affective Disorders subscale. This finding illustrates three related points: Experimental feigners are comparable to real-world feigners; feigners are prone to overgeneralizing when reporting symptoms; and the accuracy of the subscales of the SIMS is at best equal to its total score.

A study by Rogers et al. (2005) suggests that subscales might be useful to discriminate between malingering and factitious disorder. In their simulation experiment, these authors instructed undergraduates to role-play individuals who feign symptoms for financial reasons or individuals who feign for intrapsychic reasons (e.g., medical attention). Those with the first set of instructions (malingering condition) were found to score particularly high on the Neurological Impairment subscale (N), whereas those with the second set of instructions (factitious condition) scored especially high on the Affective Disorders subscale (AF). Rogers et al. (2005) concluded that the AF minus N index might provide a basis for differentiating between the two types of feigning. Obviously, this interesting finding needs independent replication.

Another promising development is the attempt by Rogers and co-workers (2014) to design new SIMS subscales that are based on the established detection strategies of rare symptoms (RS) and unlikely symptom combinations (SC). Rogers and colleagues employed a simulation design in an inpatient sample and found the RS and SC scales to produce very large[1] effect sizes: Cohen's ds of 2.0 and 1.6, respectively.

Aggregated data

An analysis of the effect of language on SIMS scores revealed that Dutch non-clinical controls score lower than English, German, and Spanish nonclinical controls (weighted means of 4.7 and 7.6, respectively; $p < .001$, two tailed; $d = -.8$). We have no plausible explanation for this finding, but it means that data generated by Dutch non-clinical controls might not generalize well to other cultures. In all other groups, language effects are absent (ps $> .05$). Therefore, the results of these groups were pooled in Table 6. This table aggregates the data of 41 studies—comprising 125 subsamples and 4810 SIMS protocols[2]—in which mean SIMS scores are reported. Table 6 shows weighted means, standard deviations, Cohen's ds when various samples are compared to the pooled scores of honest patients, claimants, and defendants, the corresponding 95% confidence intervals for ds, and the overlap of score distributions that the ds imply. This overlap is an upper bound estimate of the diagnostic accuracy of the SIMS: The smaller the overlap between scores of feigners and honest patients, the higher the potential diagnostic accuracy.

It can be argued that studies involving experimental feigners are of little use because such feigners have other motives than real-world feigners. Moreover, the outcomes of experimental feigning research depend highly on the specific procedures employed in a study, such as quality and elaboration of the scenario with which participants are instructed (Nies & Sweet, 1994; Merckelbach, Smeets, & Jelicic, 2009). Still, experimental feigners might serve as a proxy for feigners in a real-world situation. For instance, Brennan and Gouvier (2006) and Clegg et al. (2009) contrasted experimental feigners with feigning claimants. Neither study found significant differences between the SIMS scores of both groups. A similar correspondence between SIMS scores of experimental and real-life feigners was evident in the preliminary findings of Santamaría Fernández (2013). The data displayed in Table 6 corroborate the overlap between SIMS scores of experimental feigners and at least some categories of real-world feigners. As can be seen, SIMS scores of experimental feigners fall within a range that is typical for feigning defendants.

In contrast to experimental feigners, experimental honest responders (i.e., nonclinical controls) are poor substitutes for their real-world counterparts:[3] Table 6 shows that

---

[1]We followed the recommendation of Rogers et al. (2005) for the qualification of effect sizes: moderate ($\geq$ .75), large ($\geq$ 1.25), and very large ($\geq$ 1.50).

[2]The data summarized in Table 6 are based on 4562 individuals who—due to within-participant (i.e., test–retest) designs—contributed 4810 SIMS protocols. The studies used are marked with * in the references.

[3]Note that, in clinical practice, the SIMS will never be administered to nonclinical honest responders, because they, by definition, do not claim to suffer from psychopathology.

Table 6. Weighted means, standard deviations, Cohen's ds[a], corresponding 95% confidence intervals, and overlap between distributions[b] of Structured Inventory of Malingered Symptomatology (SIMS) scores in various samples, based on 41 studies, including 125 subsamples and 4810 SIMS protocols

Population	Number of samples	Total sample size	Weighted mean SIMS raw score	95% Confidence interval of weighted mean	Weighted standard deviation	Cohen's d[a]	95% Confidence interval of Cohen's d	Overlap[b]
Nonclinical controls (English, German, Spanish)[f]	12	648	7.6	5.9–9.2	4.1	−.9	−1.2 to −.6	48%
Nonclinical controls (Dutch)[f]	19[d]	688[e]	4.7	3.5–5.9	3.4	−1.2	−1.5 to −.9	38%
Honest patients[c]	17	742	16.1	13.4–18.9	8.8	N/A	N/A	N/A
Honest claimants	6	563	13.5	11.9–15.1	6.3	N/A	N/A	N/A
Honest defendants	3	141	13.2	10.9–15.5	8.4	N/A	N/A	N/A
Feigning claimants	9	238	23.7	20.4–27.0	8.7	1.1	0.8 to 1.4	41%
Feigning defendants	3	102	38.2	33.6–42.8	14.6	2.8	2.4 to 3.2	9%
Experimental feigners (Psychiatric)	17	615	35.8	31.3–40.3	12.3	2.5	2.1 to 2.9	12%
Experimental feigners (Cognitive)	15	526	34.5	29.8–39.2	12.2	2.4	2.0 to 2.8	13%
Experimental feigners (Miscellaneous)	10	212	29.6	25.1–34.1	10.1	1.8	1.5 to 2.1	23%
Experimental feigners (Clinical knowledge)	6	171	35.5	31.3–39.7	11.8	2.5	2.1 to 2.9	12%
Experimental feigners (Forewarned of SVT)	4	71	23.3	18.4–28.2	10.7	1.0	.7 to 1.3	45%
Experimental feigners (Knowledge + forewarned)	4	93	26.5	22.1–30.9	10.1	1.4	1.1 to 1.7	32%

The data summarized in this table are based on 4562 individuals who—due to within-participant (i.e., test–retest) designs—contributed 4810 SIMS protocols. The studies used are marked with * in the references. [a]Calculated using the weighted mean score of honest patients, claimants, and defendants. [b]Overlap between the distributions of scores of the specified populations and the scores of honest patients, claimants, and defendants. [c]The honesty of the vast majority of these patients is assumed, but not established by symptom validity testing. [d]Including 1 within-participants (i.e., test–retest) sample. [e]Including 24 within-participants (i.e., test–retest) SIMS protocols. [f]Aside from the significant difference ($p < .001$, two tailed. Cohen's $d = −.8$) between Dutch nonclinical controls and English, German, and Spanish nonclinical controls, there were no significant differences between the mean scores of Dutch, English, German, and Spanish respondents to the SIMS.

honest claimants, defendants, and patients score higher than nonclinical controls. The effect size of genuine psychopathology on SIMS scores[4] is moderate ($d = .9$ in English, German, and Spanish samples; $d = 1.2$ in Dutch samples). Thus, as stated earlier, studies that rely on nonclinical controls overestimate the specificity and positive predictive power of the SIMS.

Table 6 also reveals that feigning claimants score significantly lower than feigning defendants ($d = 1.1$ vs. $d = 2.8$). The overlap between the distributions of SIMS scores of honest and feigning claimants is 41%, whereas the overlap between scores of honest and feigning defendants is only 9%. This suggests that the diagnostic accuracy of the SIMS is better in criminal law settings than in civil law settings. Feigning defendants usually aim for reduced criminal responsibility, which arguably requires more radical symptoms than the typical disability that is needed for the compensation that claimants commonly seek.

Another conclusion that can be drawn from Table 6 is that experimental feigners who feign psychiatric symptoms produce total SIMS scores that are similar to those of experimental feigners who fake cognitive deficits ($d = 2.5$ vs. $d = 2.4$; overlap of 12% and 13%, respectively). This implies that the sensitivity of the SIMS to feigned cognitive deficits is roughly equivalent to its sensitivity to feigned psychiatric symptoms.

In the majority of experimental studies, participants were explicitly asked to feign certain types of symptoms (i.e., either cognitive or psychiatric symptoms). Samples in which it was left to the participants to decide which type of psychopathology they would attempt to feign are brought together in Table 6 under the caption "Experimental feigners (Miscellaneous)". The greater freedom to choose which type of symptoms to feign has a mitigating effect on SIMS scores of experimental feigners: It reduces the effect of experimental feigning (d) from 2.5 to 1.8.

Furthermore, Table 6 confirms that SIMS scores of experimental feigners can be subdued considerably by forewarning of symptom validity testing (d drops from 2.5 to 1.0), but not by relevant knowledge about psychopathology (d remains 2.5). However large the effect of forewarning, it still falls short of enabling coached feigners to produce scores that fall completely within the range of honest claimants and patients. More specifically, forewarning increases the overlap between distributions of SIMS scores of experimental feigners and honest patients from approximately 13% to 45%. Also evident from Table 6 is that the effects of forewarning and relevant clinical knowledge do not interact to produce a greater moderating effect on SIMS scores of feigners. As mentioned before, clinical knowledge curbs the effect of forewarning on the sensitivity of the SIMS.

DISCUSSION

Limitations of the SIMS

Despite its solid internal consistency and broad coverage of (bogus) psychopathology, the SIMS suffers from several limitations. These are inherent to the rationale behind its construction. Smith and Burger (1997) developed the SIMS for the purpose

---

[4]Calculated by contrasting the weighted mean SIMS score of nonclinical respondents with that of honest patients, claimants, and defendants.

of forensic screening. As a consequence, the SIMS covers a number of extreme dysfunctions (in particular with regard to claimed mental retardation and amnestic syndrome) that defendants might feign in the context of, for example, pleas of diminished criminal responsibility. However, such extremes will have lower base rates outside criminal settings. Milder, less disabling cognitive impairments are not addressed by the SIMS items. Therefore, the sensitivity of the SIMS might be limited in civil forensic assessment, where claims of moderate or even mild impairment are much more common.

A related point is that the SIMS items focus mainly on bizarre and atypical symptoms (but see below for exceptions). This capitalization on bizarre symptoms renders the SIMS readily recognizable as an SVT, (e.g., Merten et al., 2008). More importantly, the lack of alternative (i.e., other than pseudosymptom) items constrains the diagnostic possibilities of the SIMS. The absence of items that cover genuine symptoms makes it difficult to determine whether heightened SIMS scores might be potentially related to genuine psychopathology. Similarly, without control items it is not possible to identify irrelevant response patterns such as indiscriminate affirmation or rejection of items (resulting from, e.g., recalcitrance or indifference). Thus, its exclusive reliance on pseudopsychopathology prevents the SIMS from differentiating between feigning and aberrant test behavior.

The presence of genuine symptoms within SIMS subscales is a fourth shortcoming. The Affective Disorders subscale has been criticized particularly for this reason (Widder, 2011). Illustrative examples of moot items from this subscale are items #32 ("I have trouble …") and #52 ("I do not seem …"). The overlap of this subscale with genuine depressive symptoms might bias honest patients toward the cutoff. For example, Kobelt, Göbber, Bassler, and Petermann (2012) found that the prevalence of SIMS failure (i.e., SIMS total score > 16) among patients with depression was disproportionally high relative to other clinical groups (57% vs. 4–24%).

There are other SIMS items that might tap into genuine psychopathology, notably items #15 (memory problems), #20 (head injury), and #44 (tinnitus). Furthermore, Widder (2011) opined that items #5 (changed taste of food), #6 (laughing rarely), #10 (changing body shape), #43 (difficulties maintaining sleep), and #66 (being inactive) might reflect genuine psychopathology.

Another source of measurement error might be the wording of some items: In particular item #65 ("When I hear …"), which contains a double negation, is difficult to understand for some persons. As well, item #21 ("There are six …") is strictly speaking logically correct reasoning, although endorsing it would count as an indication for feigning. This item is of the "there-are-living-100-people-in-the-US" type. One could imagine an autistic patient with a tendency toward concretism endorsing this item. A better way of formulating this type of item would be, "There are *exactly* 100 people living in the US".

An asset of the SIMS is its low comprehension difficulty; it does not require a high reading level (i.e., Flesch-Kincaid Scale 5.3 suffices; Smith, 2008). Nevertheless, in a sample of nonclinical controls ($N = 100$), SIMS scores were found to be slightly dependent on verbal intelligence ($B = 0.11$, $R^2 = .18$, $p < .05$), but not on age, sex, or education (although the dependence on education was significant when verbal intelligence was not controlled for; Giger & Merten, 2013). Indeed, the SIMS has yielded low specificity rates among individuals with intellectual disability and care is warranted

in making determinations of feigning in this group based on the traditional cutoffs. It could be that individuals with intellectual disability produce heightened SIMS scores because of their diminished capacity to comprehend SIMS items. However, it might also be that low intelligence predisposes individuals to engage in more transparent forms of feigning (e.g., Solomon et al., 2010).

A limitation of the literature on the SIMS is the lack of test–retest reliability data. The stability of SIMS scores over time has been assessed only in small, nonclinical control samples. Merckelbach and Smith (2003) gave the Dutch version of the SIMS twice to 24 female undergraduate students, using a 3-week test interval. They obtained a test–retest correlation coefficient of .72, which can be regarded as satisfactory. Cima and colleagues (2003) administered the German version of the SIMS to 18 undergraduates (14 men, 4 women) twice, with a 6-week interval in between, and found an outstanding test–retest correlation: .97.

Suboptimal specificity

When the traditional cut scores (i.e., > 14 and > 16) are employed, the SIMS meets several requirements for a sound SVT (Hartman, 2002): Its hit rate (sensitivity) is acceptable; it is sensitive to differential prevalence; and it is robust against coaching. These qualities generalize across gender (e.g., Alwes, Clark, Berry, & Granacher, 2008; Wisdom et al., 2010), race (Edens et al., 2007; Vitacco et al., 2007), and language (i.e., Dutch: Merckelbach & Smith, 2003; German: Cima et al., 2003; and Spanish: González Ordi & Santamaría Fernández, 2009). In addition, the SIMS is relatively easy to administer and interpret, and it measures a wide range of symptoms that are likely targets for feigning (see Dandachi-FitzGerald & Merckelbach, 2013).

However, the ability of the original SIMS cutoffs to classify honest responders with psychopathology correctly is not satisfactory: Cut scores of > 14 and > 16 have generally yielded low specificity rates in honest patient samples. In patients with schizophrenia, individuals with intellectual disability, and patients suffering from psychogenic non-epileptic seizures, the SIMS generates many positive results (Benge et al., 2012; Graue et al., 2007; Peters et al., 2013). Certain characteristics of schizophrenia, such as deficits in reality monitoring (Radaelli, Benedetti, Cavallaro, Colombo, & Smeraldi, 2013) lack of illness insight (Shad, Tamminga, Cullum, Haas, & Keshaven, 2006), and cognitive impairment (Schaefer, Giangrande, Weinberger, & Dickinson, 2013; but see Stevens et al., 2014), may predispose patients to produce high SIMS scores. Similarly, intellectual disability may contribute to heightened SIMS scores due to, for example, deficiencies in verbal comprehension, abstract thinking, and judgment (American Psychiatric Association, 2013). Psychogenic disorders such as psychogenic movement disorder and non-epileptic seizures (PNES) may involve cognitive biases (e.g., "jumping to conclusions"; Pareés, Kassavetis, et al., 2012) and abnormalities in attention and perception (Pareés, Saifee, et al., 2012) that might lead patients to experience and report peculiar symptoms. This might explain why these patients sometimes fail on PVTs (Drane et al., 2006; Heintz et al., 2013) and SVTs such as the SIMS (Benge et al., 2012).

Although specific deficits inherent to schizophrenia, intellectual disability, and psychogenic disorder can put patients at risk of generating raised SIMS scores, high

SIMS scores should not be dismissed lightly. Studies that employed the SIMS in these populations (Benge et al., 2012; Graue et al., 2007; Peters et al., 2013) did not conform to known-groups designs, which complicates the interpretation of their results. The authors of these studies noted that their patients were not involved in legal proceedings and were not bent on gaining benefits in any way. Yet, obvious external incentives are not a prerequisite for dubious scores on SVTs (e.g., Fox, 2011). Also, absence of evidence for external incentives should not be taken as evidence for the absence of such incentives. In fact, research into *hidden agendas* of patients drives home the point that a substantial portion of patients (up to 42%) have covert motives for obtaining secondary gains associated with their patient status (e.g., financial support, help or attention from others, stimulant medication, work or study related privileges, or evasion of responsibilities; van Egmond & Kummeling, 2002; van Egmond, Kummeling, & van Balkom, 2005).

Clinicians should be circumspect in explaining high SIMS scores as a "cry for help": Such an interpretation might be valid, but it requires evidence that is independent from methods that rely on self-report. Furthermore, there is some evidence that feigning psychopathology is associated with reduced rather than increased treatment intensity, which runs counter to the notion of "cry for help" (Greene, 1988).

Given that the SIMS is (mainly) composed of non-genuine symptoms, it stands to reason that prominent scores (e.g., > 21) imply that respondents are either unwilling or unable to report their symptoms accurately, which means that diagnostic follow-up examinations are warranted. Just as the honesty of patients should not be discredited solely on the basis of elevated SIMS scores, so excessively raised SIMS scores should not be dismissed only because salient external incentives are absent and/or the patient has a diagnosis of, say, schizophrenia. Not only would such an approach ignore factitious motives to feign psychopathology (e.g., sympathy, attention, and care that come with the "sick role"), it would also disregard the possibility that patients may engage in feigning or exaggeration of symptoms.

Besides genuine inability and feigning, there are many reasons why patients might score above the SIMS cutoff. For example, patients might respond arbitrarily to a test in an attempt to obstruct the assessment or in order to be done with it as quickly as possible. Or they might fail to comprehend complex sentences (e.g., item #13 "There is nothing …"). Obviously, obstruction, rashness, and insufficient language comprehension bear no direct relation to feigning. Thus, the inference that a heightened SIMS score represents feigning is a *secondary* clinical inference; the *primary* clinical inference is "noncredible symptom report". This primary inference means that *other* test scores and self-reports of the patient cannot be accepted at face value *either*. Indeed, a recurrent finding in the literature is that heightened SIMS scores explain a considerable part of other test scores (Dandachi-FitzGerald, Ponds, Peters, & Merckelbach, 2011; Merten, Friedel, & Stevens, 2007; van Beilen, Griffioen, Gross, & Leenders, 2009).

Optimal cutoff

In light of the considerable consequences of discrediting genuine psychopathology, the conclusion that patients, claimants, or defendants feign symptoms should only be drawn when there is solid evidence. Surely, any diagnostic decision criterion that

lacks specificity (i.e., that carries a significant false-positive risk) will have difficulties generating such evidence. The SIMS has poor specificity when cut scores of > 14 and > 16 are employed (see also Rogers et al., 2014). How problematic this is depends on the purpose for which the SIMS is utilized. When it is used as a global screening measure of symptom validity such that scores exceeding the cutoff lead to conclusive follow-up testing, suboptimal specificity is defensible. If, on the other hand, the SIMS is employed as part of a multi-method SVT battery that is utilized for conclusive assessment of feigned psychopathology, then the substandard specificity associated with cut scores of > 14 or > 16 is perilous. In that case, a cut score of > 19 (Clegg et al., 2009) or—when even more diagnostic certainty is required—> 24 (Wisdom et al., 2010) would be more appropriate. Of course, gains in specificity due to raising the cutoff come at the cost of sensitivity. Thus, setting the cutoff at > 19 or > 24 will result in safer yet fewer identifications of feigned psychopathology, such that the SIMS will only identify the most blatant forms of feigning, and hence lose its quality as a screening instrument.

Another approach would be to use a cutoff of > 19 and define a zone of less certainty as to individual classifications. Table 7 shows a tentative taxonomy. A similar procedure has been proposed by Rogers et al. (1992) for the SIRS. Its advantage is that a strict and artificial dichotomy between feigning and honest responding is avoided.

Thus, the optimal cutoff varies depending on the rationale for using the SIMS (i.e., screening versus conclusive assessment). The trade-off between the power to detect feigned psychopathology (i.e., sensitivity) and the ability to avoid false positives (i.e., specificity) forces clinicians to decide in advance whether specificity should take precedence over sensitivity. If the goal is to screen for possible cases of invalid symptom reporting, then a cut score of > 16 suffices. If, however, it is crucial to avoid false positives (as would be the case in many instances of clinical or forensic assessment), then a cut score of at least > 19 is called for. In populations with particularly heightened SIMS scores due to genuine psychopathology (e.g., schizophrenia, intellectual

Table 7. Possible cut scores of the Structured Inventory of Malingered Symptomatology (SIMS) and corresponding interpretations and recommendations as to their use

Cut score	Considerations
>16	Recommended when the SIMS is employed as a *screen* for feigned psychopathology. Carefully investigate and possibly exclude false-positive classifications.
>19	Recommended when the SIMS is employed as part of a test battery that is utilized for conclusive assessment of feigned psychopathology. It yields lower sensitivity, but higher specificity (reduced risk of false-positive classification).
>16 – >19	Combined cutoffs. Use scores from 17 to 19 as indicating *possible* feigning, or *relatively mild* feigning. Follow-up testing is warranted.
>24	Only recommended when the SIMS is employed as part of a test battery for conclusive assessment in populations with particularly heightened SIMS scores due to *genuine* psychopathology (e.g., schizophrenia, intellectual disability). It yields high specificity, but low sensitivity (high risk of false-negative classification).
General Caveat	Heightened SIMS scores do not necessarily reflect feigned psychopathology: They might also be the result of irrelevant responding due to, for example, fatigue, frustration, indifference, defiance, or incomprehension.

disability), cut scores of > 19 for screening and > 24 for conclusive assessment would be highly recommendable. In sum, selection of a particular cutoff should be considered carefully and in advance: The data in Table 6 and the taxonomy in Table 7 might be helpful in this regard.

Multi-method approach

In general, people who feign psychopathology tend to overgeneralize: They might feign both psychiatric symptoms *and* cognitive deficits (i.e., they might over-report symptoms *and* underperform on cognitive tests). Moreover, feigners of psychopathology overgeneralize to cognitive impairments more frequently than the other way around (i.e., over-reporters underperform more often than underperformers over-report; Alwes et al., 2008; see also Green, Rosenfeld, Belfi, Rohlehr, & Pierson, 2012; Heinze & Purisch, 2001). However, people might also be highly selective in the symptoms or impairments they feign. Astute or informed feigners will even limit their complaints to specific signs and symptoms of a particular disorder or disability. It is therefore essential that symptom validity assessment includes multiple measures covering diverse domains of symptomatology during various stages of the evaluation (Boone, 2009; Heilbronner et al., 2009).

Even though the SIMS taps into several areas of psychopathology, it is advisable to combine it with other SVTs, preferably PVTs, which tap underperformance, a dimension that is relatively independent of symptom over-reporting as indexed by the SIMS (Nelson, Sweet, Berry, Bryant, & Granacher, 2007; Ruocco et al., 2008; see also Table 5). Precisely because PVTs aim at another dimension of symptom validity testing, they are susceptible to other error sources than the SIMS (Dandachi-FitzGerald et al., 2011; Green et al., 2012).

Using the SIMS in conjunction with other validity tests allows for a significant reduction in false-positive risk; namely, through adherence to the rule that a respondent has to fail at least *two* validity tests in order to be classified as a feigner (Giger et al., 2010). Clinicians and researchers who decide to employ the SIMS are well-advised to bear this rule, also known as the *two-failure rule* (Victor et al., 2009), in mind. Concerns that the efficacy of the two-failure rule diminishes when more than two validity tests are administered (Berthelson, Mulchan, Odland, Miller, & Mittenberg, 2013) are dispelled by recent empirical findings (Davis & Millis, 2014; Larrabee, 2014).

Because its administration is simple and brief, the SIMS is a fair candidate for inclusion in a multi-method approach to symptom validity assessment. The original cut scores of the SIMS yield substandard specificity, which renders them inappropriate for confirming suspicious scores on other validity tests. On the other hand, the sensitivity of these cutoffs is relatively high. This makes them more suitable to rule feigning *out*, rather than rule it *in*. Hence, when using its original cutoffs, the SIMS is best used as first instrument in a multi-method approach. An advantage of using the SIMS as primary screen is that its subscales may provide an indication of the type of psychopathology a respondent might feign, which could facilitate the selection of validity tests that are tailored to specific (pseudo)psychopathology.

Alternatives

Further proliferation of the SIMS comes with two risks. The first is that information about its cutoffs may become so prevalent that well-informed and calculating respondents will adapt their test behavior in order to pass the SIMS. It is true that the SIMS is presently robust against coaching. However, there might come a time when potential respondents can educate themselves (e.g., through the Internet; see Bauer & McCaffrey, 2006; Ruiz, Drake, Glass, Marcotte, & van Gorp, 2002) on how to feign psychiatric complaints while evading detection when confronted with the SIMS (for a recent example, see the Wikipedia entry on the TOMM).[5]

A second risk is that widespread use of the SIMS can lead to a form of thoughtless routine on the part of clinicians. Sociologists have termed this phenomenon the *performance paradox* (van Thiel & Leeuw, 2002). It arises when examinees have learned to mask inferior performance by adjusting their response to performance indicators such that they obtain superior test results (Meyer & Gupta, 1994). A similar scenario might unfold with instruments such as the SIMS. That is, respondents may come to learn how to obscure a sophisticated form of feigning by scoring reassuringly low on the SIMS.

Thus, the adequate sensitivity of the SIMS with cut scores of up to > 16 is not likely to last forever. Bearing this in mind, it is sensible to consider alternatives to the SIMS. One candidate is the M Test (Beaber, Marston, Michelli, & Mills, 1985; for reviews, see Boone, 2013, or Smith, 2008). The M Test is a 25-item self-report measure with a dichotomous response format that contains genuine symptoms of schizophrenia as well as absurd beliefs and bogus symptoms. The M Test focuses on feigned schizophrenia and is hence ill-suited to assess other types of feigned psychopathology. Nevertheless, the M Test might be valuable in a forensic context where blatant forms of feigning can be expected. In a direct comparison of their sensitivities, the M Test outperformed the SIMS in a sample of defendants suspected of feigning incompetence to stand trial, with detection rates being .93 and .87, respectively (Heinze & Purisch, 2001). However, in samples where the psychopathology that respondents attempt to feign is relatively mild, the diagnostic qualities of the M Test are less satisfactory, especially with regard to specificity (Boone, 2013).

Another alternative is the Miller-Forensic Assessment of Symptoms Test (M-FAST; Miller, 2001; see also Smith, 2008). The M-FAST is a 25-item structured interview that (much like the vastly more extensive SIRS) probes into unusual and implausible symptomatology. In a known-groups comparison—with the SIRS serving as external criterion—of patients involved in competency to stand trial evaluations, the M-FAST held its ground as firmly as the SIMS did, producing very large effect sizes (*d* M-FAST: 2.7, *d* SIMS: 3.1; Vitacco et al., 2007). Similar effects were attained in a sample of claimants of personal injury or workers' compensation (*d* M-FAST: 3.0, *d* SIMS: 2.6; Alwes et al., 2008). The M-FAST might be a good alternative to the SIMS when patients or defendants have reading difficulties.

Of course, symptom validity assessment is not limited to freestanding screening measures such as the SIMS and the M-FAST. Obvious alternatives for comprehensive assessment of symptom validity include the SIRS, but also clinical inventories with

---

[5]http://en.wikipedia.org/wiki/Test_of_Memory_Malingering (URL of August 2014).

embedded symptom validity scales (for a qualitative review, see Boone, 2013), such as the Minnesota Multiphasic Personality Inventory-2 Restructured Form (MMPI-2-RF; Ben-Porath & Tellegen, 2008) and the Personality Assessment Inventory (PAI; Morey, 2007). A notable advantage of the SIMS over these more extensive instruments lies in its short administration time. As said before, however, the traditional cut scores of the SIMS are more suited to rule feigning *out* than to rule it *in*, while the cutoffs of these more comprehensive tools are geared toward the opposite.

Conclusion

In accordance with previous research, we found the SIMS to be a fairly sensitive test that is reasonably robust against coaching. A serious weakness of the SIMS is its poor specificity when the original cut scores (> 14 and > 16) are employed. These cut-offs are more effective in ruling feigning out than ruling it in, which is consistent with the SIMS's status as a screen for feigned psychopathology. The substandard specificity of the original cut scores of the SIMS can be worked around by combining the SIMS with PVTs and other SVTs, and by raising the cut score, although the latter solution sacrifices sensitivity for specificity. Furthermore, we advise clinicians to adopt a position of respectful skepticism toward respondents with a heightened SIMS score. Although an assumption of honesty might be naïve in, for example, a forensic context, the burden of proof for the conclusion that an individual feigns symptoms rests on the shoulders of diagnostic experts. A deviant SIMS score alone does not meet the burden of proof, but it should be an impetus for follow-up investigation.

REFERENCES

References marked with an asterisk indicate studies included in Table 6.

* 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.). Arlington, VA: American Psychiatric Publishing.

Bauer, L., & McCaffrey, R. J. (2006). Coverage of the Test of Memory Malingering, Victoria Symptom Validity Test, and Word Memory Test on the Internet: Is test security threatened? *Archives of Clinical Neuropsychology, 21*, 121–126.

Beaber, R. J., Marston, A., Michelli, J., & Mills, M. J. (1985). A brief test for measuring malingering in schizophrenic individuals. *American Journal of Psychiatry, 142*, 1478–81.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J. E., & Erbaugh, J. K. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4*, 561–571.

* 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*, 439–444.

Ben-Porath, Y. S., & Tellegen, A. (2008). *The Minnesota Multiphasic Personality Inventory-2 Restructured Form: Manual for administration, scoring, and interpretation*. Minneapolis, MN: University of Minnesota.

Berthelson, L., Mulchan, S. S., Odland, A. P., Miller, L. J., & Mittenberg, W. (2013). False positive diagnosis of malingering due to the use of multiple effort tests. *Brain Injury, 27*, 909–916.

Boone, K. B. (2007). A reconsideration of the Slick et al (1999) criteria for malingered neurocognitive dysfunction. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 29–49). New York, NY: Guilford.

Boone, K. B. (2009). The need for continuous and comprehensive sampling of effort/response bias during neuropsychological examinations. *The Clinical Neuropsychologist, 23*, 729–741.

Boone, K. B. (2013). *Clinical practice of forensic neuropsychology: An evidence-based approach*. New York, NY: Guilford.

Brennan, A. M., & Gouvier, W. D. (2006). Are we genuinely studying malingering? A profile and comparison of feigned and suspected malingerers. *Applied Neuropsychology, 13*, 1–11.

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

Capilla Ramírez, P., González Ordi, H., & Santamaría Fernández, P. (2008). Detteción de simulación en dolor crónico. In H. González Ordi & P. Santamaría Fernández (Eds.), *Inventario Estructorado de Simulación de Síntomas – SIMS Manual [Structured Inventory of Malingered Symptomatology – SIMS Manual]* (pp. 72–77). Madrid: Tea Ediciones.

* Chen, D. K., Izadyar, S., Collins, R. L., Benge, J. F., LeMaire, A. W., & Hrachovy, R. A. (2011). Induction of psychogenic nonepileptic events: Success rate influenced by prior induction exposure, ictal semiology, and psychological profiles. *Epilepsia, 52*, 1063–1070.

* Chen, D. K., Maheshwari, A., Franks, R., Trolley, G. C., Robinson, J. S., & Hrachovy, R. A. (2014). Brief group psychoeducation for psychogenic nonepileptic seizures: A neurologist-initiated program in an epilepsy center. *Epilepsia, 55*, 156–166.

* Cima, M., Hollnack, S., Kremer, K., Knauer, E., Schellbach-Matties, R., Klein, B., & Merckelbach, H. (2003). Strukturierter Fragebogen Simulierter Symptome [Structured Inventory of Malingered Symptomatology]. *Nervenarzt, 74*, 977–986.

* Cima, M., & van Oorsouw, K. (2013). The relationship between psychopathy and crime-related amnesia. *International Journal of Law and Psychiatry, 36*, 23–29.

* 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*, 239–254.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York, NY: Routledge.

* 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.

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*, 812–828.

Davis, J. J., & Millis, S. R. (2014). Examination of performance validity test failure in relation to number of tests administered. *The Clinical Neuropsychologist, 28*, 199–214.

Derogatis, L. R. (1994). *SCL-90-R: Administration, scoring and procedures manual*. Minneapolis, MN: National Computer Systems.

Drane, D. L., Williamson, D. J., Stroup, E. S., Holmes, M. D., Jung, M., Koerner, E., … Miller, J. W. (2006). Cognitive impairment is not equal in patients with epileptic and psychogenic nonepileptic seizures. *Epilepsia, 47*, 1879–1886.

* 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.

Fox, D. D. (2011). Symptom validity test failure indicates invalidity of neuropsychological tests. *The Clinical Neuropsychologist, 25*, 488–495.

Freeman, T., Powell, M., & Kimbrell, T. (2008). Measuring symptom exaggeration in veterans with chronic posttraumatic stress disorder. *Psychiatry Research, 158*, 374–380.

* Geraerts, E., Jelicic, M., & Merckelbach, H. (2006). Symptom over-reporting and recovered memories of childhood sexual abuse. *Law and Human Behavior, 30*, 621–630.

* Giger, P., & Merten, T. (2013). Swiss population-based reference data for six symptom validity tests. *Clínica y Salud, 24*, 153–159.

* 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.

* 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.

* González Ordi, H., & Santamaría Fernández, P. (2008). Detection of malingering in clinical, medicolegal, and forensic settings. In H. González Ordi & P. Santamaría Fernández (Eds.), *Inventario Estructorado de Simulación de Síntomas – SIMS Manual [Structured Inventory of Malingered Symptomatology – SIMS Manual]* (pp. 60–66). Madrid: Tea Ediciones.

González Ordi, H., & Santamaría Fernández, P. (2009). *Inventario Estructorado de Simulación de Síntomas – SIMS Manual [Structured Inventory of Malingered Symptomatology – SIMS Manual]*. Madrid: Tea Ediciones.

* González Ordi, H., Santamaría Fernández, P., Blasco Saiz, J.L., & Pallardó Durá, L. (2008). Estudio de patron de exageración de síntomas en contextos médico-legales [Study concerning the pattern of symptom exaggeration in medicolegal contexts]. In H. González Ordi & P. Santamaría Fernández (Eds.), *Inventario Estructorado de Simulación de Síntomas – SIMS Manual [Structured Inventory of Malingered Symptomatology – SIMS Manual]* (pp. 69–71). Madrid: Tea Ediciones.

* González Ordi, H., Santamaría Fernández, P., & Fernández Marín, P. (2010). Precisión predictive del Inventario Estructurado de Simulación de Síntomas – SIMS en el contexto medicolegal [Predictive accuracy of the Structured Inventory of Malingered Symptomatology – SIMS in the medicolegal setting]. *Edupsykhé, 9*, 3–22.

* González Ordi, H., Santamaría Fernández, P., & Matalobos Veiga, B. (2008). Inventario estructurado de simulatión de síntomas – SIMS. Una medida de screening de simulación de síntomas psicopatológicos y neuropsicológicos [Structured Inventory of Malingered Symptomatology – SIMS. A screening measure for feigned psychopathological and neuropsychological symptoms]. In H. González Ordi & P. Santamaría Fernández (Eds.), *Inventario Estructorado de Simulación de Síntomas – SIMS Manual [Structured Inventory of Malingered Symptomatology – SIMS Manual]* (pp. 54–60). Madrid: Tea Ediciones.

* 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.

Green, D., & Rosenfeld, B. (2011). Evaluating the gold standard: A review and meta-analysis of the Structured Interview of Reported Symptoms. *Psychological Assessment, 23*, 95–107.

Green, D., Rosenfeld, B., Belfi, B., Rohlehr, L., & Pierson, A. (2012). Use of measures of cognitive effort and feigned psychiatric symptoms with pretrial forensic psychiatric patients. *International Journal of Forensic Mental Health, 11*, 181–190.

Green, P. (2003). *Green's Word Memory Test for Windows: User's manual*. Edmonton, Canada: Green's Publishing.

Greene, R. L. (1988). The relative efficacy of F-K and the obvious and subtle scales to detect overreporting of psychopathology on the MMPI. *Journal of Clinical Psychology, 44*, 152–159.

Hartman, D. E. (2002). The unexamined lie is a lie worth fibbing: Neuropsychological malingering and the Word Memory Test. *Archives of Clinical Neuropsychology, 17*, 709–714.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Participants, Conference (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.

Heintz, C. E. J., van Tricht, M. J., van der Salm, S. M. A., van Rootselaar, A. F., Cath, D., Schmand, B., & Tijssen, M. A. J. (2013). Neuropsychological profile of psychogenic jerky movement disorders: Importance of evaluating noncredible cognitive performance and psychopathology. *Journal of Neurology, Neurosurgery & Psychiatry, 84*, 862–867.

* 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*, 23–52.

Inman, T. H., Vickery, C., Berry, D., Lamb, D., Edwards, C., & Smith, G. (1998). Development and initial validation of a new procedure for evaluating adequacy of effort given during neuropsychological testing: The Letter Memory Test. *Psychological Assessment, 10*, 120–127.

* 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 feigners. *Journal of Psychopathology and Behavior Assessment, 28*, 19–22.

* Jelicic, M., Merckelbach, H., Candel, I., & Geraerts, E. (2007). Detection of feigned cognitive dysfunction using special malinger 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. (2007). 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.

Kobelt, A., Göbber, J., Bassler, M., & Petermann, F. (2012). Beschwerdenvalidität im Rahmen stationärer psychosomatischer Rehabilitation [Symptom validity in psychosomatic rehabilitation]. *Rehabilitation, 51*, 349–355.

Kunst, M., Winkel, F. W., & Bogaerts, S. (2011). Recalled peritraumatic reactions, self-reported PTSD, and the impact of malingering and fantasy proneness in victims of interpersonal violence who have applied for state compensation. *Journal of Interpersonal Violence, 26*, 2186–2210.

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. (2014). False-positive rates associated with the use of multiple performance and symptom validity tests. *Archives of Clinical Neuropsychology, 29*, 364–373.

* 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.

Lilienfeld, S. O., & Andrews, B. P. (1996). Development and preliminary validation of a self-report measure of psychopathic personality traits in noncriminal populations. *Journal of Personality Assessment, 66*, 488–524.

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*, 437–455.

* Merckelbach, H., & Collaris, J. (2012). Mother Theresa doesn't help here: Lack of moral priming effects on malingered symptom reports and what we can learn from it. *Psychologica Belgica, 52*, 201–215.

* Merckelbach, H., Jelicic, M., & Pieters, M. (2011). The residual effect of feigning: How intentional faking may evolve into a less conscious form of symptom reporting. *Journal of Clinical and Experimental Neuropsychology, 33*, 131–139.

Merckelbach, H., Smeets, T., & Jelicic, M. (2009). Experimental feigning: Type of malingering scenario makes a difference. *The Journal of Forensic Psychiatry & Psychology, 20*, 378–386.

* 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., Diederich, C., & Stevens, A. (2008). Vorgetäuschte Beschwerden nach Distorsionstrauma der Halswirbelsäule: Eine experimentelle Feigningsstudie [Feigned symptoms after whiplash injury: An experimental feigning study]. *Aktuelle Neurologie, 35*, 8–15.

* Merten, T., Friedel, E., Mehren, G., & Stevens, A. (2007). Über die Validität von Persönlichkeitsprofilen in der nervenärztlichen Begutachtung [On the validity of personality profiles in psychiatric assessment]. *Nervenarzt, 78*, 511–520.

Merten, T., Friedel, E., & Stevens, A. (2007). Die Authentizität der Beschwerdenschilderung in der neurologisch-psychiatrischen Begutachtung: eine Untersuchung mit dem Strukturierten Fragebogen Simulierter Symptome [Authenticity of symptom report in independent neurological and psychiatric examinations: A study with the Structured Inventory of Malingered Symptomatology]. *Praxis der Rechtspsychologie, 17*, 140–154.

* Merten, T., Lorenz, R., & Schlatow, S. (2010). Posttraumatic Stress Disorder can easily be faked, but faking can be detected in most cases. *German Journal of Psychiatry, 13*, 140–149.

Merten, T., & Merckelbach, H. (2013). Symptom validity testing in somatoform and dissociative disorders: A critical review. *Psychological Injury and Law, 6*, 122–137.

Merten, T., Thies, E., Schneider, K., & Stevens, A. (2009). Symptom validity testing in claimants with alleged posttraumatic stress disorder: Comparing the Morel Emotional Numbing Test, the Structured Inventory of Malingered Symptomatology, and the Word Memory Test. *Psychological Injury and Law, 2*, 284–293.

Meyer, M. W., & Gupta, V. (1994). The performance paradox. *Research in Organizational Behavior, 16*, 309–369.

Miller, H. A. (2001). *M-FAST: Miller-Forensic Assessment of Symptoms Test professional manual*. Odessa, FL: Psychological Assessment Resources.

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

Morey, L. C. (2007). *The Personality Assessment Inventory professional manual* (2nd ed.). Odessa, FL: Psychological Assessment Resources.

Nelson, N. W., Sweet, J. J., Berry, D. T. R., Bryant, F. B., & Granacher, R. P. (2007). Response validity in forensic neuropsychology: Exploratory factor analytic evidence of distinct cognitive and psychological constructs. *Journal of the International Neuropsychological Society, 13*, 440–449.

Nies, K. J., & Sweet, J. J. (1994). Neuropsychological assessment and malingering: A critical review of past and present strategies. *Archives of Clinical Neuropsychology, 9*, 501–552.

Pareés, I., Kassavetis, P., Saifee, T. A., Sadnicka, A., Bhatia, K. P., Fotopoulou, A., & Edwards, M. J. (2012). Jumping to conclusions' bias in functional movement disorders. *Journal of Neurology, Neurosurgery & Psychiatry, 83*, 460–463.

Pareés, I., Saifee, T. A., Kassavetis, P., Kojovic, M., Rubio-Agusti, I., Rothwell, J. C., & … Edwards, M. J. (2012). Believing is perceiving: Mismatch between self-report and actigraphy in psychogenic tremor. *Brain, 135*, 117–123.

* Peters, M. J. V., Jelicic, M., Moritz, S., Hausschildt, M., & Jelinek, L. (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*, 64–77.

Plohmann, A. M., & Merten, T. (2013). The Third European Symposium on Symptom Validity Assessment – Facts and controversies. *Clínica y Salud, 24*, 197–203.

Radaelli, D., Benedetti, F., Cavallaro, R., Colombo, C., & Smeraldi, E. (2013). The reality monitoring deficit as a common neuropsychological correlate of schizophrenic and affective psychosis. *Behavioral Sciences, 3*, 244–252.

Rogers, R., Bagby, R. M., & Dickens, S. E. (1992). *The Structured Interview of Reported Symptoms (SIRS)*. 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*, 244–257.

* 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*, 21–41.

* Rogers, R., Robinson, E. V., & Gillard, N. D. (2014). The SIMS screen for feigned mental disorders: The development of detection-based scales. *Behavioral Sciences and the Law, 32*, 455–466.

Ruiz, M. A., Drake, E. B., Glass, A., Marcotte, D., & van Gorp, W. G. (2002). Trying to beat the system: Misuse of the Internet to assist in avoiding the detection of psychological symptom dissimilation. *Professional Psychology: Research and Practice, 33*, 294–299.

Ruocco, A. C., Swirsky-Sacchetti, T., Chute, D. L., Mandel, S., Platek, S. M., & Zillmer, E. A. (2008). Distinguishing between neuropsychological malingering and exaggerated psychiatric symptoms in a neuropsychological setting. *The Clinical Neuropsychologist, 22*, 547–564.

Santamaría Fernández, P. (2013, June). *The Structured Inventory of Malingered Symptomatology (SIMS) in civil forensic evaluations*. Paper presented at the Third European Symposium on Symptom Validity Assessment, Würzburg, Germany.

Schaefer, J., Giangrande, E., Weinberger, D. R., & Dickinson, D. (2013). The global cognitive impairment in schizophrenia: Consistent over decades and around the world. *Schizophrenia Research, 150*, 42–50.

Schagen, S., Schmand, B., de Sterke, S., & Lindeboom, J. (1997). Amsterdam Short-Term Memory Test: A new procedure for the detection of feigned memory deficits. *Journal of Clinical and Experimental Neuropsychology, 19*, 43–51.

Shad, M. U., Tamminga, C. A., Cullum, M., Haas, G. L., & Keshavan, M. S. (2006). Insight and frontal cortical function in schizophrenia: A review. *Schizophrenia Research, 86*, 54–70.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*, 107–120.

Slick, D. J., Hopp, G., Strauss, E., & Spellacy, F. J. (1996). Victoria Symptom Validity Test: Efficiency for detecting feigned memory impairment and relationship to neuropsychological tests and MMPI-2 Validity Scales. *Journal of Clinical and Experimental Neuropsychology, 18*, 911–922.

Slick, D. J., Sherman, E. M. S., & Iverson, I. (1999). Diagnostic criteria for malingered neurocog-nitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuro-psychologist, 13*, 545–561.

Smith, G. P. (2008). Brief screening measures for the detection of feigned psychopathology. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (3rd ed.). (pp. 323–339). New York: Guilford.

* Smith, G. P., & Burger, G. K. (1997). Detection of malingering: Validation of the Structured Inventory of Malingered Symptomatology (SIMS). *Journal of the American Academy on Psychiatry and Law, 25*, 180–183.

Solomon, R. E., Boone, K. B., Miora, D., Skidmore, S., Cottingham, M., Victor, T., & … Zeller, M. (2010). Use of the WAIS-III Picture Completion Subtest as an embedded measure of response bias. *The Clinical Neuropsychologist, 24*, 1243–1256.

Stevens, A., Schneider, K., Liske, B., Hermle, L., Huber, H., & Hetzel, G. (2014). Is subnormal cognitive performance in schizophrenia due to lack of effort or to cognitive impairment? *German Journal of Psychiatry, 17*, 1–9.

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

* van Beilen, M., Griffioen, B. T., Gross, A., & Leenders, K. L. (2009). Psychological assessment of malingering in psychogenic neurological disorders and non-psychogenic neurological disorders: Relationship to psychopathology levels. *European Journal of Neurology, 16*, 1118–1123.

van Egmond, J., & Kummeling, I. (2002). A blind spot for secondary gain affecting treatment outcome. *European Psychiatry, 17*, 46–54.

van Egmond, J., Kummeling, I., & van Balkom, T. (2005). Secondary gain as hidden motive for getting psychiatric treatment. *European Psychiatry, 20*, 416–421.

van Thiel, S., & Leeuw, F. L. (2002). The performance paradox in the public sector. *Public Performance & Management Review, 25*, 267–281.

Victor, T. L., Boone, K. B., Serpa, J. G., Buehler, J., & Ziegler, E. A. (2009). Interpreting the mean-ing of multiple symptom validity test failure. *The Clinical Neuropsychologist, 23*, 297–313.

* 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.

* Vossler-Thies, E., Stevens, A., Engel, R. R., & Licha, C. (2013). Erfassung negativer Antwortverzerrungen mit der deutschen Fassung des "Personality Assessment Inventory", dem "Verhaltens- und Erlebensinventar" [Capturing negative response distortion with the German version of the Personality Assessment Inventory]. *Diagnostica, 59*, 73–85.

Widder, B. (2011). Beurteilung der Beschwerdenvalidität [Evaluation of symptom validity]. In B. Widder, & P. W. Gaidzik (Eds.), *Begutachtung in der Neurologie* (2nd ed., pp. 64–92). Stuttgart: Thieme.

Widows, M. R., & Smith, G. P. (2005). *Structured Inventory of Malingered Symptomatology professional manual*. Odessa, FL: Psychological Assessment Resources.

* 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*, 118–125.

* Zimmermann, P., Kowalski, J. T., Alliger-Horn, C., Danker-Hopfe, H., Engers, A., Meermann, R., & Hellweg, R. (2013). Detection of malingering in the assessment of occupational disability in the military. *German Journal of Psychiatry, 16*, 54–60.