Examining Base Rates of Symptom Endorsement and the Roles of Sex and Depressive Symptoms on the Structured Inventory of Malingered Symptomology (SIMS) in a Non-clinical Population

Aqsa Zahid[1] · Sonya Dhillon[2] · Konstantine K. Zakzanis[1,2]

Abstract

Symptom validity tests (SVTs) assist in ensuring that our diagnostic conclusions are grounded in credible psychometric results. The Structured Inventory of Malingered Symptomology (SIMS) is often employed to identify overreporting among respondents. The present study examined base rate scores on the SIMS in an undergraduate sample and the role of biological sex and the presence of depressive symptoms. Participants ($N = 93$) completed the SIMS, the Beck Depression Inventory (BDI-II), and the Personality Assessment Inventory (PAI). Of the participants, 50.00% were characterized as overreporting based on the cutoff score suggested in the SIMS manual (> 14); 40.00% were characterized as such based on a well-suggested cutoff score in the current literature (> 16); and 8.57% were characterized as such when the highest suggested cutoff score in the literature was applied (> 24). The multivariate result was not significant for sex or language status. A strong positive correlation was found between the SIMS total score with the BDI-II total score ($r = .425$), the PAI depression-cognitive scale ($r = .502$) and the PAI depression-affective scale ($r = .491$). The SIMS total score must be considered on an individual basis to decrease the possibility of a false positive conclusion regarding symptom veracity. Depressive symptomology may be a key moderator in that cognitive biases held by individuals with depressive symptoms may be associated with elevated scores on the SIMS.

Keywords Malingering · Depressive symptoms · Symptom validity · SIMS · Type I error · Healthy sample

Introduction

The practice of clinical psychology is rooted in empiricism by using existing, rigorous, and the best available evidence to inform practice. In the context of psychological assessment, practice is informed by the use of valid and reliable psychological test measures. Results of psychological assessments inform diagnostic decision making, prognostic opinion, treatment choice, and may assist the trier of fact where decisions regarding disability status and compensatory awards are necessary. Such endeavors are dependent, however, on the accuracy of responses from those put before

us in a clinical setting. Hence, it is essential that accurate information from a patient is garnered whereby the breadth and severity of symptoms are reported in relatively true form, free of minimization, exaggeration, fabrication and/or denial of symptoms (Bush et al., 2014). To these ends, there exists an extensive body of research literature that has detailed such response styles. Minimization refers to the downplaying or underreporting of the extent to which clinical symptoms impede on one's daily functioning (Graham et al., 2003, p. 571). Exaggeration refers to the act of overstating the nature of one's situation and/or symptoms (Graham et al., 2003, p. 229). Fabrication or overreporting is defined as the act of concocting the reality of their symptoms by intentionally overreporting them (Graham et al., 2003, p. 229). Denial of symptoms refers to the intentional neglection of unpleasant symptoms (Graham et al., 2003, p. 571).

The need to assess symptom validity in the context of an assessment is based on various factors. Valid test results allow psychologists to analyze the findings in relation to the construct of interest (Bush et al., 2014). Invalid test results

✉ Konstantine K. Zakzanis
konstantine.zakzanis@utoronto.ca

[1] Department of Psychology, University of Toronto Scarborough, Toronto, ON, Canada

[2] Graduate Department of Psychological Clinical Science, University of Toronto Scarborough, Toronto, ON, Canada

are associated with minimal confidence in the process of assessment analysis, such that conclusions drawn from these results cannot be made with confidence (Bush et al., 2014; Edens et al., 2007; van Impelen, 2018, p. 8). Within the context of an assessment, self-report measures can be used in order to assess the validity of symptom reporting (Malcore et al., 2015). Indeed, in the last few decades, there has been an increase in the scientific research and knowledge of symptom validity, and thus, symptom validity tests (SVTs) were created (Dandachi-FitzGerald et al., 2013). SVTs have allowed psychologists to recognize non-credible symptomology to a reliable manner in the context of claims regarding injury and work-related disability (Dandachi-FitzGerald et al., 2013). One such commonly employed SVT measure in the context of neuropsychological assessment is the Structured Inventory of Malingered Symptomology (SIMS; Smith & Burger, 1997).

Structured Inventory of Malingered Symptomology (SIMS; Smith & Burger, 1997)

The SIMS (Smith & Burger, 1997) is a 75-item forced-choice screening measure to detect fabricated or exaggerated symptoms associated with psychiatric disturbance of cognitive dysfunction. The SIMS consists of five subscales: psychosis (*P*), neurologic impairment (*NI*), amnesic disorders (*AM*), low intelligence (*LI*), and affective disorders (*AF*; Malcore et al., 2015; Widows & Smith, 2005). Each subscale consists of 15 items using a rarely reported or atypical symptom endorsement strategy. Specifically, the subscales assess the degree to which a respondent encompasses unusual psychotic symptoms (*P*; cutoff > 1), atypical neurological symptoms (*NI*; cutoff > 2), atypical memory impairment symptoms (*AM*; cutoff > 2), fabricated intelligence deficits (*LI*; cutoff > 2), and inconsistent symptoms of depression and anxiety (*AF*; cutoff > 5; Smith & Burger, 1997). Various atypical symptoms are described in the items in order to determine the validity of symptomology reported by the respondent. The quantity of "existing" symptoms is added in order to calculate a total SIMS score allowing for the assessment of general malingering presentation of a respondent (Widows & Smith, 2005). Researchers have also developed a brief version of the SIMS (Malcore et al., 2015) which requires less time to complete in comparison to the original SIMS while also maintaining the overall utility of the measure.

The extant research literature is varying regarding which specific cutoff score may represent valid and invalid symptom reporting (van Impelen, 2018, p. 13). In comparison to the subscales, the SIMS total score is likely to be more accurate in differentiating genuine and feigning patients (Widows & Smith, 2005). The total score has a high alpha coefficient (0.82) and sufficient test–retest reliability ($r = 0.72$; Widows

& Smith, 2005), and thus, researchers often work to determine which cutoff score provides the most accuracy in the classification of symptom reporting as valid or invalid. To do so, researchers may examine the positive predictive power (PPP; the probability that an individual with a score above the cutoff score is malingering) and the negative predictive power (NPP; the likelihood an individual scoring below the cutoff score is being honest in their responses). The PPP and NPP are important values to estimate a cutoff value that successfully categorizes individuals based on the score they obtain on a measure. A higher value in power indicates better likelihood that an individual is correctly categorized. For example, a PPP score of 0.90 suggests that there is a 90% chance that an individual who scored about the cutoff value has been accurately categorized as malingering; a NPP score of 0.68 suggests that there is only a 28% chance that the symptom presentation of an individual who scores below the cutoff score is really valid. Thus, researchers aim to determine a cutoff value with both high PPP and NPP values.

When regarding the SIMS as a screening tool, the cutoff score of > 14 is recommended as not only has it been suggested in the SIMS manual (Smith & Burger, 1997), but it carries the best overall discrimination with a misclassification rate of 10.3% in simulated malingerers and 15.4% in clinical malingerers (Poythress et al., 2001). However, other researchers have shown that a cutoff score of > 16 is relatively more accurate in diagnostic decision making (Merckelbach & Smith, 2003; van Impelen et al., 2014). This value has a PPP of 0.90 and a NPP of 0.98 using a low base rate of 19% (Merckelbach & Smith, 2003). Other researchers have recommended a higher total cutoff value of > 19 to increase specificity (Clegg et al., 2009). Moreover, and further complicating the interpretability of cut scores, Wisdom and colleagues (2010) suggested a much higher cutoff score of > 24 which was derived from a forensic sample. This cutoff score has been suggested to only be utilized when the SIMS is administered in a test battery to populations where higher scores are expected due to psychopathology (van Impelen et al., 2014).

Overall, the SIMS yields high sensitivity with both > 14 and > 16 cutoff score criteria. As such, it is suggested that both cutoff scores are acceptable in ruling out non-credible reporting (van Impelen et al., 2014). The specificity of the SIMS is improved when coupled with other SVTs and PVTs, along with an increase in the SIMS cutoff score. However, increasing the cutoff score may result in a trade-off between sensitivity for specificity. It is important to note that non-clinical individuals tend to have significantly lower scores on the SIMS compared to patient controls, suggesting that the SIMS is sensitive to genuine psychopathology (van Impelen et al., 2014). To this end, a higher base rate value increases the PPP, whereas a lower base rate leads to a larger NPP (van Impelen et al., 2014). For both > 14 and > 16 cutoff scores,

effective NPP is obtained, such that there is a high probability that an examinee is being forthright in their symptom report. As base rates increase, the NPP decreases especially for a cutoff score of > 14 (van Impelen et al., 2014). The probability of a responder feigning their symptoms if their SIMS score is above the cutoff score depends on the base rate of feigning in their particular population. The PPP for > 14 and > 16 cutoff scores are low and decrease drastically as base rates decrease. In populations where feigning is uncommon, false-positive identification is of higher probability. Hence and again, this suggests that the SIMS is sensitive to genuine psychopathology and may overestimate feigning in patients (van Impelen et al., 2014). With regard to the highest cutoff score suggested to date, > 24, there is a higher probability of a false negative conclusion due to the low sensitivity and high specificity.

With respect to depressive symptomology, it is possible that the presence of such symptoms may moderate scores on the SIMS. As evidenced among a veteran sample (Miskey et al., 2019), failure on the SIMS was positively associated with a posttraumatic stress disorder diagnosis, negatively associated with distress tolerance, and positively associated with the severity of symptoms. Other researchers have drawn associations between other disorders and invalid SIMS scores (e.g., somatoform disorders; Merten & Merckelbach, 2013). Moreover, researchers have investigated whether there exist biological sex differences in scores obtained on the SIMS. This area of research is quite limited; however, Merten and colleagues (2020) recently examined such differences in an in-patient sample and found that females endorsed more items on the SIMS relative to males. In one other study, Wiedmaier (2011) examined sex differences on the SIMS in a non-clinical German sample and also found that females endorsed more items than males. More research is needed in this area to confirm and possibly replicate these findings. Overall, the base rates of symptom endorsement on the SIMS in non-clinical samples who have no history of psychological disorder are crucial to examine. Additional variance that might exist in this population secondary to sex differences and depressive symptomology are unknown.

The Present Study

The purpose of the present study was to examine base rate scores on the SIMS in healthy individuals with no history of psychopathology. We further sought to determine if sex differences and depressive symptomatology were associated with total and subscale scores on the SIMS. Specifically, we set out to address three hypotheses:

1. We set out to determine if healthy individuals endorse and score above the cutoff score on the SIMS. Here, we hypothesized that a considerable number of individuals will score above the cutoff scores > 14 and > 16 on the SIMS; the false positive rate is expected to be relatively higher when the more conservative cutoff value (i.e., > 14) is applied.

2. We set out to examine the association between SIMS scores and the sex of participants. We hypothesized that female participants will score significantly higher than males on the SIMS, as found in previous studies which examined sex differences on the SIMS (Merten et al., 2020; Wiedmaier, 2011).

3. Finally, we set out to examine whether depressive symptoms were related to SIMS scores. We hypothesized that there will be a positive association between total scores on the SIMS and the BDI-II which is congruent with our previous findings wherein we found that participants exert poor effort on performance validity measures secondary to attentional and encoding demands (An et al., 2012).

Methods

Participants

A total of 93 participants were recruited for this study. This sample consisted of undergraduate students enrolled in an introductory psychology class at the University of Toronto and were recruited through the Department of Psychology's Experimental Participation System. All participants received course credit in exchange for participation. Participants were excluded if they were less than 18 years of age, scored above the cutoff values on the Personality Assessment Inventory (PAI) inconsistency scale (ICN), the infrequency scale (INF), and the positive impression scale (PIM), or had a history of psychological disorder and/or neurological injury; this information was confirmed via a self-report demographics survey.

Measures

The SIMS (Smith & Burger, 1997) is a 75-item forced-choice instrument in which participants are instructed to read a statement and either agree or disagree with it. This test assesses malingered psychopathology as well as neuropsychological symptoms. As suggested by Merckelbach and Smith (2003), the recommended SIMS total cutoff score is >16. The sensitivity of this score is 0.93 and the specificity is 0.98 (Merckelbach & Smith, 2003). However, a total cutoff score of > 14 is mentioned in the SIMS manual (Smith & Burger, 1997).

The Beck Depression Inventory (BDI-II; Beck et al., 1997) is a 21-item self-report measure used to assess depression-related symptoms among respondents. This questionnaire was administered in order to understand the relationship between scores on the BDI-II and the SIMS.

The PAI (Morey, 2007) is a 344-item self-report questionnaire assessing psychopathology in adults. This measure consists of 22 non-overlapping scales: four validity scales, 11 clinical scales, five treatment scales, and two interpersonal scales. The PAI was included to assess each participants ability to attend to all items and respond consistently, based on their scores on the four validity scales: the ICN, the INF, the PIM, and the negative impression scale (NIM).

A post-experimental questionnaire was employed to determine if instructions provided at the outset of the study were understood and the measures were completed to the best of their ability. A demographics questionnaire was also administered to collect the demographics information for each participant including age, years of education starting from and including grade one, biological sex, mental health history, etc.

Procedure

The current study was a post hoc analysis of a larger study in which data was collected with multiple experimental groups. For the purposes of the current study, only the data from the control group were used in the analysis. The study protocols were approved by the University of Toronto Office of Research Ethics.

Data collection was conducted online via Qualtrics software. Informed consent was obtained prior to the completion of any testing measures. Participants were instructed to attend to all items with full honesty and accuracy. The demographics questionnaire, BDI-II, SIMS, PAI, and a post-experimental questionnaire were all administered in the same order for all participants through a secure link. Upon completion of all test measures, participants read and acknowledged receipt of a debrief form, and were compensated with course credit.

Data Analysis

Statistical analyses were utilized to determine whether or not there exists a significant difference in scores on the SIMS, PAI, and BDI-II between male and female participants. Specifically, a multivariate analysis of variance (MANOVA) was conducted to examine if there were mean level differences between males and females on dependent variables (e.g., BDI-II total score, SIMS subscale and total scores, and PAI validity scale scores). Moreover, and to examine if there were group differences between individuals who are native and non-native English language speakers on the dependent variables, an additional MANOVA was conducted. To control for multiple comparisons, Bonferroni correction was applied. As there were 11 comparisons made in each set of MANOVA's, the critical p value was set at $p = 0.005$. Pearson correlation coefficients were computed

to examine the association among variables of interests, namely the BDI-II, the SIMS, the PAI NIM scale, and the PAI depression cognitive (DEP-C), affective (DEP-A), and physiological (DEP-P) scales. The correlation coefficients would build on the understanding of the relationship between symptom validity and depression. Moreover, to understand whether such an association is relevant to only the SIMS or all SVTs in general, correlation analyses among the PAI NIM and the BDI-II were conducted.

Results

Sample Characteristics

In this study, 93 participants were recruited. Based on the exclusion criteria, 16 participants were excluded due to self-reporting a psychological disorder, five participants were excluded due to self-reporting a neurological injury, and two participants were excluded due to over-reporting on the PAI validity scales. Thus, the final sample consisted of 70 participants. Of these participants, 29 were male and 39 were female; two participants preferred not to self-report their biological sex. The average age of the participants was 19.81 years ($SD = 1.73$; range = 18–26 years). The mean years of education was 13.70 years ($SD = 1.83$). The ethnicity of the sample was comprised of the following self-reported groups: East Asian (40.0%), South Asian (30.0%), Other (14.3%), Black (8.6%), and Southeast Asian (5.7%). Moreover, 44.28% of the sample were native English speakers.

BDI-II

On the BDI-II, the mean total score obtained among all participants was 12.99 ($SD = 9.81$). This falls in the minimal range of depression on the BDI-II.

PAI

On the PAI, all but two participants [who were excluded from the study] scored below 73 T, 68 T, 92 T, and 68 T on the ICN, INF, NIM, and PIM subscales, respectively. The mean T score on the PIM scale was 50.03 ($SD = 10.09$); the mean T score on the INF scale was 50.01 ($SD = 10.10$); the mean T score on the NIM scale was 50.13 ($SD = 10.03$); and finally, the mean T score on the ICN scale was 49.45 ($SD = 9.15$).

SIMS

The cutoff values for each subscale and the total score on the SIMS and the false-positive rate among participants are

Table 1 False-positive rates on SIMS subscale scores based on manual cutoff values ($n = 70$)

	Cutoff score	False-positive rate (%)
NI	> 2	
Male		24.14
Female		17.95
Full sample		20
AF	> 5	
Male		37.93
Female		64.1
Full sample		51.43
P	> 1	
Male		13.79
Female		33.33
Full sample		24.29
LI	> 2	
Male		65.52
Female		84.62
Full sample		74.29
AM	> 2	
Male		20.69
Female		35.9
Full sample		28.57

NI neurologic impairment scale, *AF* affective disorders scale, *P* psychosis scale, *LI* low intelligence scale, *AM* amnesic disorders scale

shown in Tables 1 and 2, respectively. The mean NI subscale score was 2.17 ($SD = 2.00$); the mean was above the cutoff of > 2 for the *NI* subscale, as 20.00% of participants scores exceeded this cutoff. The mean obtained for this sample was also above the cutoff of > 5 for the AF subscale ($M = 6.18$; $SD = 2.63$) and 51.43% of the sample scored above 5. The mean P subscale score was 1.12 ($SD = 1.76$). The mean was above the cutoff of > 1 for this subscale and 24.29% of the sample scored above the cutoff value. The mean LI subscale score was 4.73 ($SD = 1.87$). The mean was above the cutoff of > 2 and 74.29% of the participants scored above this cutoff value. Finally, the mean was above

Table 2 False-positive rates on SIMS total score based on various cutoff values ($n = 70$)

	False-positive rate with > 14 (%)	False-positive rate with > 16 (%)	False-positive rate with > 24 (%)
Total score			
Male	34.48	24.14	3.45
Female	64.1	53.85	12.82
Full sample	50	40	8.57

the cutoff of > 2 on the AM subscale ($M = 2.12$; $SD = 1.67$), as 28.57% of the participants scored above the cutoff value.

With respect to the cutoff values provided in the literature, the SIMS total score false positive rate was calculated for the major cutoff scores provided (refer to Table 2). Among the full sample, the mean total score on the SIMS was 15.49 ($SD = 8.62$). A cutoff score of > 14 has been stated in the SIMS manual; this cutoff value had a false positive rate of 50.00%. In terms of the highly suggested cutoff in the literature of > 16, a false positive rate of 40.00% was found. Finally, based on the highest suggested cutoff value in the literature (> 24; Wisdom et al., 2010) a false positive rate of 8.57% was found.

Group Level Analyses

The means, standard deviations, and results of the MANOVA for biological sex and language status are displayed in Table 3. The significance level for the MANOVA was set at $p < 0.005$, after Bonferroni correction for multiple comparisons. The multivariate result was not significant for sex, Pillai's Trace = 0.467, $F = 1.83$, df = (11, 23), $p = 0.107$, indicating no difference in the level of the dependent variables between males and females. On the BDI-II, the multivariate result was not significant for sex, $F(1, 33) = 3.78$; $p = 0.051$. The multivariate result was not significant for sex on the PAI ICN subscale, $F(1, 33) = 2.72$; $p = 0.109$), the PAI INF subscale, $F(1, 33) = 0.295$; $p = 0.591$), the PAI NIM subscale, $F(1, 33) = 0.004$; $p = 0.953$), and the PAI PIM subscale, $F(1, 33) = 0.356$; $p = 0.555$), indicating no difference in the level of the dependent variables between males and females on the PAI validity scales. The multivariate result was not significant for sex on the SIMS AF subscale, $F(1, 33) = 4.45$; $p = 0.043$), the SIMS LI subscale, $F(1, 33) = 1.67$; $p = 0.205$), the SIMS AM subscale, $F(1, 33) = 2.28$; $p = 0.141$), the SIMS P subscale, $F(1, 33) = 2.00$; $p = 0.167$), the SIMS NI subscale, $F(1, 33) = 0.117$; $p = 0.734$), and the SIMS total score, $F(1, 33) = 3.78$; $p = 0.060$), indicating no difference in the SIMS subscale scores and total score between males and females.

The multivariate result was not significant for language status, Pillai's Trace = 0.445, $F = 1.68$, $df = (11, 23)$, $p = 0.142$, indicating no difference in the level of the dependent variables between native and non-native English language speakers. On the BDI-II, the multivariate result was not significant for language status, $F(1, 33) = 0.722$; $p = 0.402$). Similarly, multivariate result was not significant for language status on the PAI ICN subscale, $F(1, 33) = 0.032$; $p = 0.859$), the PAI INF subscale, $F(1, 33) = 2.93$; $p = 0.097$), the PAI NIM subscale, $F(1, 33) = 0.437$; $p = 0.513$), and the PAI PIM subscale, $F(1, 33) = 0.000$; $p = 0.992$), indicating no difference in the PAI validity scales between native and non-native English language speakers. Finally, the

Table 3 Means, standard deviations, and estimates of group difference for biological sex and language status on variables of interest

| | Biological sex | | | | Language status | | | |
| | Male ($n=29$) | Female ($n=39$) | | | Native English ($n=31$) | Non-native English ($n=39$) | | |
	M (SD)	M (SD)	F	p	M (SD)	M (SD)	F	p
SIMS								
NI	2.04 (1.70)	2.27 (2.22)	.117	.734	2.35 (2.56)	2.03 (1.45)	1.34	.255
AF	5.22 (2.80)	6.82 (2.34)	4.45	.043	5.84 (2.87)	6.44 (2.45)	.034	.854
P	.71 (1.49)	1.40 (1.90)	2.00	.167	1.33 (2.27)	.94 (1.19)	.000	1.00
LI	4.05 (1.66)	5.15 (1.89)	1.67	.205	5.22 (2.28)	4.38 (1.45)	4.61	.039
AM	1.64 (1.22)	2.46 (1.87)	2.28	.141	2.19 (1.72)	2.06 (1.65)	.447	.509
Total	12.67 (7.05)	17.82 (8.84)	3.78	.060	16.27 (10.78)	14.86 (6.45)	.057	.814
BDI-II								
Total	9.67 (8.36)	15.56 (10.08)	3.78	.051	14.43 (10.74)	11.81 (8.98)	.722	.402
PAI								
ICN	46.22 (7.38)	51.43 (9.66)	2.72	.109	48.72 (7.88)	50.01 (10.10)	.032	.859
INF	50.17 (10.10)	49.92 (10.25)	.295	.591	48.90 (10.99)	50.82 (9.50)	2.93	.097
NIM	50.36 (12.45)	50.00 (10.88)	.004	.953	51.91 (11.80)	48.71 (9.28)	.437	.513
PIM	50.01 (10.88)	50.04 (9.78)	.356	.555	50.77 (10.29)	49.43 (10.05)	.000	.992

N sample size, *M* mean, *SD* standard deviation, *SIMS* Structured Inventory of Malingered Symptomology, *NI* neurologic impairment scale, *AF* affective disorders scale, *P* psychosis scale, *LI* low intelligence scale, *AM* amnesic disorders scale, *BDI-II* Beck Depression Inventory-II, *PAI* Personality Assessment Inventory, *ICN* inconsistency scale, *INF* infrequency scale, *NIM* negative impression scale, *PIM* positive impression scale

*Significance was set at $p < .004$, after Bonferroni correction

multivariate result was not significant for language status on the SIMS AF subscale, $F(1, 33)=0.034$; $p=0.854$), the SIMS LI subscale, $F(1, 33)=4.61$; $p=0.039$), the SIMS AM subscale, $F(1, 33)=0.447$; $p=0.509$), the SIMS P subscale, $F(1, 33)=0.000$; $p=1.00$), the SIMS NI subscale, $F(1, 33)=1.34$; $p=0.255$), and the SIMS total score, $F(1, 33)=0.057$; $p=0.814$), indicating no difference in SIMS scores between native and non-native English language speakers.

Correlation Analyses

With respect to correlation analyses, a strong positive correlation between BDI-II scores and SIMS total scores was found ($r=0.425$, $p<0.001$). There was a strong positive correlation between the PAI NIM *T* scores and the BDI-II total scores ($r=0.439$, $p<0.001$). Moreover, the SIMS total score correlated with the PAI depression cognitive (DEP-C) scale ($r=0.502$, $p<0.001$), the PAI depression affective (DEP-A) scale ($r=0.491$, $p<0.001$), and the PAI depression physiological (DEP-P) scale ($r=0.203$, $p=0.134$). Moreover, the PAI NIM scale correlated with the PAI DEP-C scale ($r=0.570$, $p<0.001$), the PAI DEP-A scale ($r=0.627$, $p<0.001$), and the PAI DEP-P scale ($r=0.446$, $p<0.001$).

Discussion

The present study examined base rate scores on the SIMS in an undergraduate sample without history of psychopathology and moreover, sought to determine the role of potential moderating variables including sex differences and symptoms of depression. In keeping with our previously published findings (An et al., 2012), we hypothesized that a considerable number of individuals will score above the cutoff of > 14 and > 16 on the SIMS and that sex differences will moderate SIMS scores. We also hypothesized that individuals who reported higher scores on the BDI-II would be more prone to produce a higher score on the SIMS, which is also congruent with our previous findings wherein we found that participants exert poor effort on performance validity measures secondary to attentional and encoding demands (An et al., 2012).

Our first hypothesis was supported as 50.00% of this random sample of 70 individuals without a documented history of psychopathology scored above the cutoff score of > 14, while 40.00% of participants scored above the total cutoff score of > 16. This suggests that both the BDI-II and the SIMS are magnified. More specifically, the high scores on the SIMS raise the concern of invalid BDI-II scores. It is important to note, however, that almost all participants scored

below the high-score cutoff values on the four validity scales in the PAI; those who scored above the cutoff values on the PAI ICN, INF, and PIM validity scales were excluded. The average PAI NIM T score was 50.13, suggesting that this sample was not over-reporting depressive symptomology. Our findings also suggest that that the current cutoff score proposed in the research literature may be problematic. There exists a need to re-consider the total score cutoff to minimize false-positive rates. Based upon our findings, utilizing a cut-off value of > 24 suggested by Wisdom et al. (2010) substantially reduced the false positive rate to 8.57%. In saying this, it is important to note that the participants recruited for this study were without any history of psychopathology.

Our second hypothesis was not supported, indicating that there were no significant mean-level differences between males and females on our outcome variables. Our findings are inconsistent with the previous literature wherein females scored relatively higher than males on the BDI-II in both healthy (Roelofs et al., 2013) and in clinical samples (Castro-Sànchez et al., 2012). This may, in part, be due to artifacts because of unequal sample sizes within the groups. Regardless, it is important for future studies to secure larger samples and seek to replicate these results.

The findings of the present study showed that there is a strong positive correlation between the SIMS and the BDI-II, as well as the SIMS and the PAI DEP-C and DEP-A subscales. This suggests that individuals with greater depressive symptomology generated false positive outcomes on the SIMS, resulting in the greater likelihood of the feigning categorization. Moreover, strong positive correlations were found between the PAI NIM and the PAI DEP-C, DEP-A, and DEP-P subscales. It was evident that this sample was generally not over-reporting their symptoms of depression as participants with invalid PAI scores by way of the ICN, INF, and PIM scales were excluded. Thus, individuals who self-report credible depressive symptomology still developed noncredible SIMS scores. It is important to further demonstrate whether depressive symptomology threatens the utility of the SIMS or other SVTs as well. A correlational analysis between the PAI NIM and BDI-II scores was conducted, resulting in a strong positive correlation ($r = 0.439$, $p < 0.001$), suggesting that the noncredible reporting was not specific to items on the SIMS and may be applicable to SVTs in general.

Here, we would assert that the presence and severity of depressive symptomology impedes with one's ability to pay attention and encode the instructions they are asked to follow (Li et al., 2016; Dhillon et al., 2020; Schatzberg et al., 2000). The current study cannot infer this directionality, however, due to the correlation design. We posit some explanations for the associations observed between BDI-II and SIMS scores in our sample. First, it is possible that those with elevated depressive symptom endorsement on the BDI-II may be exaggerating depressive symptoms, and

this influence SIMS scores; notably the SIMS is a screening device requiring additional follow up from the examiner, as would the BDI-2 in many cases. As such, the relationship among the BDI-II and SIMS scores may be related to overreporting in of itself (see Merckelbach & Smith, 2003). Another explanation is related to that of effort. While the direction of this relationship has not been established, if increases in BDI-II scores are related to increased suspiciousness of response bias according to the SIMS, this may be related to the influence and effects of inattention and truncated effort in this sample (Li et al., 2016; Schatzberg et al., 2000). Researchers have suggested that negative response biases among patients can result in elevated SIMS scores as a result of psychopathology (Merten & Merckelbach, 2013; Merten et al., 2009). To this end, individuals may struggle to perform optimally which in turn, can influence scores on the SIMS, or exhibit an overreporting response bias, which can bidirectionally influence both measures.

Individuals who scored higher on the BDI-II also scored higher on the SIMS, suggesting that depressive symptomology may be moderating SIMS total and subscale scores in healthy participants. This suggests that depression may be a potential moderating variable, resulting in a perceived non-credible total score obtained on the SIMS. Here, there exists a strong positive correlation ($r = 0.425$) between SIMS total scores and BDI-II total scores. That is, the higher one's depression score was, the greater probability existed of producing a non-credible score on the SIMS. To this end, individuals with depressive symptoms may potentially hold negative cognitive biases (Platt et al., 2015) and/or dysfunctional automatic thoughts (Gotlib & Joormann, 2010).

This is supported by Beck's Cognitive Theory of Depression which illustrates that depressed persons hold negative schemas that selectively focus on loss, worthlessness, and rejection (Gotlib & Joormann, 2010). Due to this bias in thinking, individuals tend to respond to stimuli congruent with their schema. This response is accompanied by automatic thoughts that revolve around the cognitive triad, a form of pessimistic thinking in terms of the self, the world, and the future (Gotlib & Joormann, 2010; Yousefi et al., 2016). Individuals who obtain higher scores on the BDI-II are more likely to hold negative biases as a higher BDI-II score suggests higher severity of depression (Beck et al., 1997). Hence, if the BDI-II and the SIMS are correlated, negative cognitive biases are also expected to be correlated with the SIMS. These negative tendencies in thinking may lead to the exaggeration of symptoms and may serve as one source of variance as it pertains to SIMS scores obtained by more depressed individuals. Also, factors such as attention and encoding require effort. Non-clinical participants have been found to exert low levels of effort (An et al., 2012), and thus, it is at least possible that individuals who suffer from depressive symptomology exert lower levels of effort. As

individuals with depressive symptoms present with a lack of attention and fail to effectively encode what is being asked (Li et al., 2016; Schatzberg et al., 2000), it is possible that individuals with depression who score highly on the SIMS may do so as function of their disturbed cognition.

The current study raises implications associated with forensic disability assessments. The prevalence of feigning in forensic disability and related assessments is $40 \pm 10\%$ (Larrabee et al., 2009). This rate may be overestimated as $15 \pm 15\%$ has also been suggested after careful evaluation of various prevalence rates evidenced in the literature in the context of forensic-disability assessments (Young, 2015). Nonetheless, these numbers suggest that individuals in forensic disability assessment settings may carry an external incentive to their engagement in feigning symptomology. The results of the present study suggest that many non-clinical honest responders may be characterized as feigning based on scores obtained on the SIMS when a cutoff score of > 14 or > 16 are applied. It is likely that such results may be replicated among individuals claiming a forensic disability. Specifically, it is likely that both feigning and honest respondents would be classified as noncredible reporters and they may be regarded as feigning their disability in a forensic setting. Therefore, false-positive diagnoses are more likely to be provided in the context of forensic disability assessments even if such a conclusion is falsely derived due to the use of a low cutoff score.

This directly showcases the importance of correcting the SIMS cutoff score such that individuals are correctly classified as honest or noncredible respondents. Researchers have suggested possibly increasing the cutoff score to > 19 (Clegg et al., 2009). This may be helpful in increasing the specificity and sensitivity of the SIMS in ruling in feigning and ruling out honest responders. It is also important to take into consideration the numerous variables at play in a forensic disability case in order to draw conclusions regarding symptom validity. Such variables include the injury in question, clinical records, evidence of diminished functioning, observations from clinical interviews, subject variables (e.g., age) and more (Jeffay et al., 2021). Thus, clinical judgment and careful evaluation should be exercised in addition to obtaining data from the SIMS. Additionally, the inclusion of other SVTs alongside the SIMS can also assist in obtaining more confidence in the detection of feigning symptomology in forensic disability assessments.

We are mindful of several limitations as it pertains to our study. First, the platform in which test measures were administered was not in keeping with how tests were standardized in normative samples. The reliability of the specific measures (e.g., SIMS) employed in our study when delivered by computerized test administration compared to face-to-face administration is unknown and as such, our findings should be interpreted accordingly.

Future researchers can administer the SIMS in-person and compare base rates and false-positive rates to the values found in the present study in order to develop more confidence in the current results. Moreover, the researchers could not confirm the credibility of participant responses on the BDI-II. However, as this study was part of a larger examination of symptom validity, all participants scored below the high score cutoff values for the INF, ICN, and the PIM on the PAI (Morey, 2007) suggesting that participants attended to all items and responded consistently. Nonetheless, we were unable to determine the true credibility of scores obtained on the BDI-II as many participants failed to score below the cutoff values for the SIMS, a symptom validity measure. Finally, as all participants had acceptable range scores on the NIM scale of the PAI, it is likely that there is a restricted range in response bias within our sample. Future researchers should examine a larger sample with the aim of providing a more fulsome range of responding, namely overreporting response bias and its relationship to SIMS scores. Future studies should also include community samples to improve the generalizability of the results. This will allow for a direct comparison between various SVTs to determine whether individuals are consistently classified as honest responders across these measures.

Conclusion

The present study set out to examine base rate scores on the SIMS in healthy individuals with no history of psychopathology. The results indicated that the cutoff score stated in the SIMS manual, > 14, resulted in a high false-positive rate of 50.00%. The highly suggested cutoff score of > 16 was also problematic, as this value also resulted in a high false-positive rate of 40.00%. The results also indicated that there exists a strong positive correlation between scores on the SIMS and the BDI-II, which may be since depressive symptomology may impede on an individual's attention to encode the instructions with which they are provided. These results raise implications for forensic-disability assessments, as clinicians must work to evaluate other variables apart from SVT scores such as subjective factors and interview observations.

References

An, K. Y., Zakzanis, K. K., & Joordens, S. (2012). Conducting research with non-clinical healthy undergraduates: Does effort play a role in neuropsychological test performance? *Archives of Clinical Neuropsychology, 27*, 849–857. https://doi.org/10.1093/arclin/acs085

Beck, A. T., Steer, R. A., & Brown, G. K. (1997). *Beck Depression Inventory* (2nd ed.). The Psychological Corporation.

Bush, S. S., Heilbronner, R. L., & Ruff, R. M. (2014). Psychological assessment of symptom and performance validity, response bias, and malingering: Official position of the Association for Scientific Advancement in Psychological Injury and Law. *Psychological Injury and Law, 7*, 197–205. https://doi.org/10.1007/s12207-014-9198-7

Castro-Sánchez, A. M., Matarán-Peñarrocha, G. A., López-Rodríguez, M. M., Lara-Palomo, I. C., Arendt-Nielsen, L., & Fernández-de-las-Peñas, C. (2012). Gender differences in pain severity, disability, depression, and widespread pressure pain sensitivity in patients with fibromyalgia syndrome without comorbid conditions. *Pain Medicine, 13*(12), 1639–1647. https://doi.org/10.1111/j.1526-4637.2012.01523.x

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. https://doi.org/10.1080/14789940802267760

Dandachi-FitzGerald, B., Ponds, R. W. H. M., & Merten, T. (2013). Symptom validity and neuropsychological assessment: A survey of practices and beliefs of neuropsychologists in six European countries. *Archives of Clinical Neuropsychology, 28*, 771–783. https://doi.org/10.1093/arclin/act073

Dhillon, S., Videla-Nash, G., Foussias, G., Segal, Z. V., Zakzanis, K. K. (2020) On the nature of objective and perceived cognitive impairments in depressive symptoms and real-world functioning in young adults. *Psychiatry Research,* 287112932. https://doi.org/10.1016/j.psychres.2020.112932

Edens, J. F., Poythress, N. G., & Watkins-Clay, M. M. (2007). Detection of malingering in psychiatric unit and general population prison inmates: A comparison of the PAI, SIMS, and SIRS. *Journal of Personality Assessment, 88*(1), 33–42. https://doi.org/10.1080/00223890709336832

Gotlib, I. H., & Joormann, J. (2010). Cognition and depression: Current status and future directions. *Annual Review of Clinical Psychology, 6*, 285–312. https://doi.org/10.1146/annurev.clinpsy.121208.131305

Graham, J. R., Naglieri J. A., & Weiner I. B. (Eds.) (2003). *Handbook of Psychology: Volume 10 Assessment Psychology.* New Jersey: John Wiley & Sons, Inc.

Jeffay, E., Binder, L. M., & Zakzanis, K. K. (2021). Marked intraindividual cognitive variability in a sample of healthy graduate students. *Psychological Injury and Law.* https://doi.org/10.1007/s12207-021-09417-x

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

Li, M., Zhong, N., Lu, S., Wang, G., Feng, L., & Hu, B. (2016). Cognitive behavioral performance of untreated depressed patients with mild depressive symptoms. *PLoS ONE, 11*(1), e0146356. https://doi.org/10.1371/journal.pone.0146356

Malcore, S. A., Schutte, C., Van Dyke, S. A., & Axelrod, B. N. (2015). The development of a reduced-item Structured Inventory of Malingered Symptomatology (SIMS). *Psychological Injury and Law, 8*, 95–99. https://doi.org/10.1007/s12207-015-9214-6

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*(2), 145–152. https://doi.org/10.1016/S0887-6177(01)00191-3

Merten, T., Kaminski, A., & Pfeiffer, W. (2020). Prevalence of over-reporting on symptom validity tests in a large sample of psychosomatic rehabilitation inpatients. *The Clinical Neuropsychologist, 34*(5), 1004–1024. https://doi.org/10.1080/13854046.2019.1694073

Merten, T., & Merckelbach, H. (2013). Symptom validity testing in somatoform and dissociative disorders: A critical review. *Psychological Injury and Law, 6*, 122–137. https://doi.org/10.1007/s12207-013-9155-x

Merten, T., 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. https://doi.org/10.1007/s12207-009-9057-0

Miskey, H. M., Martindale, S. L., Shura, R. D., & Taber, K. H. (2019). Distress tolerance and symptom severity as mediators of symptom validity failure in veterans with PTSD. *The Journal of Neuropsychiatry and Clinical Neuroscience, 32*(2), 161–167. https://doi.org/10.1176/appi.neuropsych.17110340

Morey, L. C. (2007). *The Personality Assessment Inventory professional manual.* Psychological Assessment Resources.

Platt, B., Murphy, S. E., & Lau, J. Y. (2015). The association between negative attention biases and symptoms of depression in a community sample of adolescents. *PeerJ, 3*, e1372. https://doi.org/10.7717/peerj.1372

Poythress, N. G., Edens, J. F., & Watkins, M. M. (2001). The relationship between psychopathic personality features and malingering symptoms of major mental illness. *Law and Human Behavior, 25*(6), 567–582. https://doi.org/10.1023/A:1012702223004

Roelofs, J., van Breukelen, G., de Graaf, L. E., Beck, A. T., Arntz, A., & Huibers, M. J. H. (2013). Norms for the Beck Depression Inventory (BDI-II) in a Large Dutch Community Sample. *Journal of Psychopathology and Behavioral Assessment, 35*, 93–98. https://doi.org/10.1007/s10862-012-9309-2

Schatzberg, A. F., Posener, J. A., DeBattista, C., Kalehzan, B. M., Rothschild, A. J., & Shear, P. K. (2000). Neuropsychological deficits in psychotic versus nonpsychotic major depression and no mental illness. *American Journal of Psychiatry, 157*(7), 1095–1100. https://doi.org/10.1176/appi.ajp.157.7.1095

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.

van Impelen, A. (2018). *All that looks grave is not grievous. Not all those who wince are in pain: studies in furtherance of validity assessment.* Maastricht University.

van Impelen, A., Merckelbach, H., Jelicic, M., & Merten, T. (2014). The Structured Inventory of Malingered Symptomatology (SIMS): A systematic review and meta-analysis. *The Clinical Neuropsychologist, 28*, 1336–1365. https://doi.org/10.1080/13854046.2014.984763

Wiedmaier, P. (2011). *Schätzung von Normalbefunden für drei Beschwerdenvalidierungstests [Estimation of normal-range results for three symptom validity tests].* VDM Verlag Dr.

Widows, M. R., & Smith, G. P. (2005). *Structured Inventory of Malingered Symptomatology professional manual.* 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*(2), 118–125. https://doi.org/10.1093/arclin/acp110

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

Yousefi, F., Redzuan, M., & Juhari, R. B. (2016). Study of comparison among information processing theory, interference model of test anxiety and Beck's cognitive theory of depression with regard to their relationship with academic achievement. *Avicenna Journal of Neuro Psycho Physiology, 3*(4), 94–100. https://doi.org/10.5812/ajnpp.58803

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