

Validation of an Abbreviated Version of the Structured Interview of Reported Symptoms in Outpatient Psychiatric and Community Settings

Debbie Green · Barry Rosenfeld · Tia Dole ·
Ekaterina Pivovarova · Patricia A. Zapf

Published online: 9 June 2007

© American Psychology-Law Society/Division 41 of the American Psychological Association 2007

Abstract This study examined the effectiveness of an abbreviated version of the Structured Interview of Reported Symptoms (SIRS-A) in identifying malingered mental illness. The SIRS-A is comprised of 69 items drawn from the SIRS (R. Rogers et al. 1992, *SIRS: Structured Interview of Reported Symptoms: Professional Manual*. Odessa, FL: Psychological Assessment Resources, Inc.), substantially reducing the administration time. A simulation design was used with three samples; 87 psychiatric outpatients who responded honestly were compared to 29 community-dwelling adults and 24 psychiatric patients instructed to malingering psychopathology. The SIRS-A generated sensitivity comparable to or exceeding that of the SIRS normative data, but specificity was poorer; many genuinely impaired patients were misclassified as malingering. Although these findings suggest the SIRS-A may be an effective means to assess malingering in psychiatric populations, further research assessing the reasons for the elevated false positive rates is necessary.

Keywords Malingering · Dissimulation · SIRS · Forensic assessment · Classification accuracy

Researchers and clinicians estimate that a notable proportion of people malingering psychopathology and/or cognitive deficits across forensic and clinical settings (e.g.,

Mittenberg et al. 2002; Rogers et al. 1998; Walters et al. 1988). Accurate assessment tools are critical in order to identify who is actually malingering. Unfortunately, many of the measures developed to identify malingered mental disorder are time consuming, limiting their practical utility (Meyer and Deitsch 1995).

The earliest and most extensively studied measure that included indices of exaggerated psychopathology is the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway and McKinley 1940) and its subsequent revision (MMPI-2; Butcher et al. 1989). Several MMPI scales have emerged as effective predictors of malingering across different samples. Examples include the Infrequency scale (*F*), the Infrequency-Psychopathology scale (*F(p)*; Arbisi and Ben-Porath 1998), the Gough Dissimulation scale (*Ds-r*; Gough 1957) and the Obvious-Subtle scales (Greene 1997; for a more detailed explanation of the various MMPI malingering scales, see Greene 1997).

In the past decade, the Structured Interview of Reported Symptoms (SIRS) has become one of the most frequently researched, used, and recommended malingering measures (Lally 2003; Miller 1999). Unlike the MMPI-2, the SIRS was developed specifically to assess whether an examinee is malingering psychological symptoms. The items of the SIRS were formulated based upon prior research of malingering tactics, and were intended to tap various strategies such as excessively endorsing rare and blatant symptoms, indiscriminately reporting symptoms, and claiming absurd or outrageous problems. The SIRS is made up of eight primary scales and five supplementary scales, as well as a number of items that are not assigned to any scale. Each of the scales are intended to assess a distinct strategy of malingering (Rogers et al. 1990, 1992).

Rogers and colleagues designed and validated the SIRS over a series of studies and clinical settings using both

D. Green · B. Rosenfeld (✉) · T. Dole ·
E. Pivovarova
Department of Psychology, Fordham University, 441 East
Fordham Road, Bronx, NY 10458, USA
e-mail: rosenfeld@fordham.edu

P. A. Zapf
John Jay College of Criminal Justice, New York, NY, USA

simulation and known-groups designs (Rogers et al. 1992). The SIRS accurately distinguished between community samples instructed to simulate mental illness and bona fide psychiatric patients (Rogers et al. 1991a). Furthermore, on each of the primary scales of the SIRS, suspected malingerers in a psychiatric hospital differed significantly from patients presumed to be responding honestly (Rogers et al. 1991). Other studies, using “antisocial individuals” instructed to mangle or individuals coached in how to evade detection while malingering, have demonstrated equally positive findings (Rogers et al. 1990, 1991). Thus, across clinical, correctional, and community samples, the primary scales of the SIRS appear valid in distinguishing malingerers from honest responders. However, suspected malingerers did not elevate the supplementary scales as had been expected. In fact, three of these scales were more highly elevated in the sample of patients with severe genuine psychopathology than they were in malingerers. Of note, because Rogers et al. (1992) contend that it is more detrimental to misclassify an honest person as malingering than it is to overlook someone faking mental illness, optimal SIRS cut-offs were established to maximize positive predictive accuracy (PPA; the proportion of individuals predicted to be malingering who actually are feigning symptoms). Their validation studies have yielded a PPA estimate of 99% whereas negative predictive accuracy (NPA; the proportion of individuals predicted to be honestly responding who are) was 64.9% with a malingering base rate of 51% (Rogers et al. 1992, p. 24). However, when the base rate of malingering is 20–25% (as is typical of criminal forensic populations), the PPA remains in the 96–97% range and the NPA increases to 86–89%. As demonstrated, PPA and NPA rates fluctuate depending upon the base rate of malingering within a particular setting (Rosenfeld et al. 1992). Therefore, sensitivity (i.e., proportion of malingerers accurately identified by the measure) and specificity rates (i.e., proportion of people accurately classified as responding honestly), which are unaffected by sample base rates, are apt to be more informative when comparing different instruments or samples than PPA and NPA. Based on published normative data, the SIRS has 99.5% specificity but only 48.5% sensitivity (based on data reported on page 24 of the test manual; Rogers et al. 1992).

Although instruments such as the MMPI-2 and the SIRS have demonstrated their utility as malingering assessment tools, their use is cumbersome. Both are exceedingly time-intensive to administer; each requiring an hour or more to complete. This constraint, coupled with the high prevalence rate of malingering in certain settings, creates a need for valid, relatively brief malingering instruments. Several measures have taken the form of either screening tools or abbreviated tests, each serving a different purpose. The intention of a screening instrument is to quickly identify

probable malingerers who can then be subjected to more intensive assessment (e.g., administered the SIRS or MMPI-2). The M Test (Beaber et al. 1985) and the Miller Forensic Assessment of Symptoms Test (M-FAST; Miller 1999) are examples of screening measures intended to identify malingered psychopathology. Although screening assessments are designed to rapidly identify possible malingerers, they do so with inherent limitations. Higher false positive rates are expected, and subsequent testing must be done to better verify response styles. In contrast, abbreviated malingering instruments are designed to replace their longer predecessors without sacrificing accuracy. For example, a number of researchers have strived to develop abbreviated versions of the MMPI (e.g., the M-168; Overall and Gomez-Mont 1974) but the ability of abbreviated versions of the MMPI to identify individuals who are malingering has rarely been studied and the limited literature that does exist has suggested poor accuracy (e.g., Ratner 1984).

There have been two prior attempts to develop briefer versions of the SIRS. Norris and May (1998) created a paper-and-pencil screening version of the SIRS that was intended to identify individuals who may be malingering. They administered 40 items that make up five of the eight primary scales to 75 inmates that had either reported psychological distress to medical personnel or had requested psychological services. Fifty participants completed the screening version, along with the original SIRS (administered as intended), with the latter used to determine whether the individual was malingering or honest. The screening version of the SIRS yielded sensitivity of 87% and specificity of 73%, with an overall classification accuracy of 80% (Norris and May 1998). However, when they extracted the screening items from the typical SIRS administration for another 25 individuals, they found a substantially higher accuracy rate (i.e., sensitivity of 100% and specificity of 90%), suggesting that their paper and pencil method of administration may have reduced validity. Unfortunately, given the lack of certainty about which were actually malingering (since the original SIRS served as the criterion against which the screening version was compared) and the paper and pencil format used in this study (deviating from the typical SIRS administration), the implications for malingering assessment are unclear.

Story (2000) also evaluated an abbreviated version of the SIRS, administering 71 items that comprise six of the SIRS scales (five primary scales and one supplementary scale). The sample consisted of 49 pre-trial forensic patients who were divided into either malingering ($n = 19$) or non-malingering ($n = 30$) groups based on the opinion of psychology staff. The abbreviated SIRS identified 15 of the 19 individuals (78.9%) thought to be malingering and correctly classified 27 of the 30 (90%) participants not believed to be malingering. The abbre-

viated SIRS was also highly correlated with a number of MMPI-2 malingering indices. Despite interesting findings, the questionable validity of the criterion groups limits the interpretation of these results.

Thus, despite the frequent need to evaluate malingering in forensic settings, there are few brief malingering instruments available to accurately determine whether an individual is feigning mental illness. This study attempted to fill this void by evaluating an abbreviated version of the SIRS (the SIRS-A). The SIRS stands as one of the most accurate malingering instruments currently available, and is therefore an optimal starting point in developing an abbreviated measure. Although two previous research groups have attempted to develop short-form versions of the SIRS, one was a self-report measure intended solely for screening purposes (Norris and May 1998), and the second used just a subset of the primary scales (Story 2000). Therefore, neither of these briefer instruments could be directly compared to the SIRS using Rogers' scoring criterion (i.e., one scale in the definite malingering range and/or three scales in the probable range; Rogers et al. 1992). Additionally, they could not assess the utility of using only the items from the eight primary scales, which are the scales that Rogers and colleagues found to have the highest psychometric properties. This study addresses both of these issues, using a simulation design in which determination of whether participants are malingering or honest is based on group assignment (rather than the SIRS) and a realistic base rate of malingering (in order to generate accurate estimates of PPA and NPA). In the first set of analyses, participants from the community were instructed to feign mental illness and were compared to bona fide psychiatric patients. This comparison evaluates the accuracy of the SIRS-A in a sample that is roughly comparable (in terms of education and socio-economic background) to individuals who might be charged with a criminal offense in contrast to genuinely impaired psychiatric patients. In the second set of analyses, the same psychiatric comparison sample was compared to another sample of psychiatric patients instructed to malingering mental illness. Previous malingering studies have primarily sampled psychologically healthy participants, instructing them to feign mental illness. Arbisi and Ben-Porath (1998) suggest that ecological validity may be improved by analyzing malingering among individuals with a genuine mental illness, in part because clinicians may have greater difficulty determining whether current symptom presentations are exaggerated or legitimate among persons with an established mental illness than among those who have no history of severe mental illness. As such, we hypothesized that psychiatric patients would be better able to feign symptoms without being detected than would non-psychiatric community members.

Method

Participants

Study participants included a sample of psychiatric outpatients ($N = 121$) with a sample of community-dwelling adults ($N = 33$). Patients were recruited from two psychiatric day-treatment programs affiliated with a state psychiatric hospital and resided either on hospital grounds or in community residences. Recruitment into the study took two forms; hospital staff provided the names of potential participants to the research team and some patients directly approached a research staff member and requested participation. All prospective participants were informed of the nature of the study, and provided written informed consent; treating clinicians reviewed each prospective participant and confirmed that they had capacity to provide informed consent. Approximately one-quarter of the patient sample ($n = 33$) were randomly assigned to the malingering condition and the remainder ($n = 88$) were instructed to complete testing honestly (see below for the instructions given). Of the 88 psychiatric patients who consented to participate and were randomized to the honest condition, one was excluded from these analyses because he was unable to complete the study measures (i.e., refused to answer many questions). Additionally, a number of patients were unable to comprehend the instructions to malingering ($n = 9$; described further below). Therefore, the final patient sample included 87 honest responders (78.4%) and 24 malingerers (21.6%), which is consistent with expected base rates of malingering.

Roughly three-quarters of the patient sample was male (75.7%, $n = 84$), with an average age of 41.6 (SD = 10.4, range: 21–66 years). The majority of the participants identified themselves as black (56.9%, $n = 62$), or Hispanic (25.7%, $n = 28$), with 15.6% Caucasian ($n = 17$), and 1.8% of mixed race ($n = 2$). The majority were diagnosed with Schizophrenia or Schizoaffective disorder (85.8%, $n = 91$); a smaller proportion was diagnosed with a mood disorder (11.3%, $n = 12$) or another Axis I disorder (2.7%, $n = 3$; diagnosis were missing for five participants). The median number of prior hospitalizations was 5.0, with a range of 0 to 100. The majority of participants had at least one prior arrest (68.6%, $n = 72$). There were no significant differences on any demographic or clinical variables between the two randomized groups (i.e., malingering versus honest) nor between those patients excluded from analysis (based on inability to comply with study instructions) and those retained ($p < .05$).

The community sample was recruited through posted fliers in a neighborhood served by the hospital (Bronx, New York). Of the 33 community participants who provided informed consent, four were excluded from data analysis be-

cause of difficulty understanding the instructions to malingering psychopathology (see below). Subsequent analyses are based on the subset of 29 participants, 17 of whom were male (58.6%). The sample ranged in age from 18 to 68 years ($M = 40.0$, $SD = 13.5$; demographic information was missing for one subject). The majority of the participants self-identified their race as black (46.4%, $n = 13$) or Hispanic (32.1%, $n = 9$), with 14.3% Caucasian ($n = 4$) and 7.1% of mixed race ($n = 2$). A number of participants reported having received some sort of mental health treatment either in the past (21.4%, $n = 6$) or at present (21.4%, $n = 6$), primarily for depression. Approximately half of the sample reported that they had been previously arrested (53.6% $n = 15$). There were no significant differences ($p < .05$) in age, education, race, psychiatric history or criminal history between those individuals who were excluded and the study participants analyzed. The community participants did not significantly differ from the patient sample in gender, age, race, or arrest history ($p < .05$). Not surprisingly, the patient sample reported a significantly higher number of previous hospitalizations than the community participants, $t(119) = 3.62$, $p < .05$.

Procedure

Three-quarters of the psychiatric participants were instructed to answer the questions honestly and to try their hardest on all tests administered. The remaining psychiatric patients and all of the community-dwelling participants were instructed to fake mental illness using the following script.

I want you to pretend that you have been arrested and charged with robbery. Your lawyer tells you that you will be convicted because someone has already identified you as the robber. You decide to pretend to be too mentally ill to face trial. Now we're going to give you a series of tests that doctors use to decide whether someone is able to stand trial, what sort of symptoms they have, and whether or not they are faking. Keep in mind that it is very important that you fool the doctors into thinking you are too mentally ill to be tried, but if they realize you are faking you may get even more time in jail so do your best without getting caught.

Participants from the three samples were administered a battery of tests consisting of the abbreviated version of the Structured Interview of Reported Symptoms (SIRS-A), along with a number of other measures that were not analyzed in this report. In order to assess understanding of the malingering instructions, participants were asked to paraphrase the instructions provided. Participants who could not accurately explain the instructions to malingering, even after additional (simplified) instructions were

provided, were eliminated from data analysis. Following completion of the study, all participants were paid \$20. This study was approved by the institutional review boards of each participating institution.

Abbreviated SIRS (SIRS-A)

The SIRS-A is comprised of the 69 items that form the eight primary scales of the SIRS (Rogers et al. 1992), and excludes all items from the supplementary scales. The eight primary scales (see Appendix) have the strongest reliability and validity (Rogers et al. 1992), and are the only indices used to assess whether a person is malingering or responding honestly. The SIRS-A retained the same ordering and wording of items as they appeared in the original SIRS, and were scored using the criteria described by Rogers and colleagues (i.e., malingering was suspected if at least one scale fell in the “definite” malingering range or three or more scales fell in the “probable” range; Rogers et al. 1992). In addition, we analyzed the predictive utility of the SIRS-A total score, calculated as the sum of all 69 items.

Statistical Analysis

Chronbach's Alpha and item-scale correlation coefficients were calculated for six of the SIRS-A scales (the Selectivity of Symptoms (SEL) and Severity of Symptoms (SEV) scales combine responses from the Blatant Symptoms (BL) and Subtle Symptoms (SU) scales, and were therefore not analyzed) and compared to those reported in the original SIRS manual. The “honest” psychiatric participants formed the comparison group against which both malingering samples were contrasted. The first set of analyses compared the sample of “honest” psychiatric patients to the “malingering” community sample. Sensitivity, specificity, positive predictive accuracy (PPA), and negative predictive accuracy (NPA) rates for the SIRS-A were examined. The chi-square statistic was used to evaluate whether the accuracy of the SIRS-A differed from the original SIRS normative data described by Rogers and colleagues (Rogers et al. 1992). In addition, Receiver Operating Characteristic (ROC) curve analyses, plotting sensitivity against 1-specificity across all possible cut-off scores, were calculated for the total number of items endorsed on the SIRS-A and the individual scales. Identical analyses compared the same control group of “honest” psychiatric patients to the “malingering” patient sample.

Results

Reliability coefficients for the SIRS-A subscales were generally comparable to those reported by Rogers et al.

(1992), with coefficient alpha exceeding .80 for all but one subscale (RO, $\alpha = .48$, see Table 1). Additionally, mean item-scale correlations on the SIRS-A scales were equal to or higher than those of the original SIRS, with a range in correlations of .36 to .64. However, a comparison of mean scale scores of the three samples (i.e., honest psychiatric patients, malingering psychiatric patients, and malingering community participants) to Rogers' published data indicated significant differences. On the Rare Symptoms (RS) and SEV scales, the patients in the current study instructed to respond honestly scored significantly higher than those sampled in the SIRS validation studies. Based on the results of independent samples t-tests, there were significant differences in mean scores on five of the individual scales between the participants instructed to malingering mental illness and Rogers' simulation groups (see Table 2).

Comparing "Honest" Psychiatric Patients to "Malingering" Community Participants

The average total score on the SIRS-A for the honest psychiatric sample was 30.16 (SD = 27.8, range: 1–97), whereas the community malingering sample endorsed significantly more SIRS-A items ($M = 101.97$, SD = 40.3, range: 5–176), $t(114) = 9.04$, $p < .001$, $d = 1.69$. Thus, the average difference in total score between the malingerers and the honest psychiatric patients was greater than 70 points. Nevertheless, psychiatric patients consistently endorsed several of the SIRS-A items (i.e., 56.4% the sample endorsed at least 20 of the items) despite instructions to respond honestly. Furthermore, three (3.4%) of the 87 psychiatric patients instructed to respond honestly scored within the "definite" malingering range on one or more of the SIRS-A scales. Nine additional patients (10.3%) obtained 3 or more scales in the "probable" malingering range. Thus, 12 of the 87 (13.8%) psychiatric patients responded to the SIRS-A in a manner that would be considered by the test authors to be indicative of malingering. However, nearly all ($n = 27$; 93.1%) of the simulators

Table 1 Alpha reliability coefficients and mean item-scale correlations for the SIRS-A and SIRS scales

Scale	Coefficient alpha		Mean item-scale correlation	
	Study	Rogers et al.	Study	Rogers et al.
RS	.80	.85	.51	.41
SC	.84	.83	.54	.35
IA	.80	.89	.54	.55
BL	.92	.92	.64	.44
SU	.89	.92	.53	.38
RO	.48	.77	.36	.24

Table 2 Mean scale score comparisons between the SIRS-A and SIRS

Scale	Honest		Malingering sample		
	Psychiatric patients		Psychiatric patients	Community	
	Study <i>M</i> (SD)	Rogers <i>M</i> (SD)	Study <i>M</i> (SD)	Study <i>M</i> (SD)	Rogers <i>M</i> (SD)
RS	2.0 (2.6) ^a	1.3 (1.7)	5.0 (4.4) ^b	8.0 (4.3)	7.8 (4.3)
SC	1.6 (2.7)	1.6 (2.3)	6.0 (5.4)	8.4 (4.5)	7.3 (4.5)
IA	.9 (1.6)	.6 (1.1)	4.0 (4.5)	5.3 (3.5)	4.7 (4.2)
BL	3.9 (4.6)	3.0 (3.9)	11.1 (8.7)	17.9 (7.4) ^b	13.7 (6.6)
SU	8.0 (6.0)	6.6 (5.8)	12.7 (8.4) ^b	19.0 (8.7)	18.0 (6.7)
SEL	8.3 (5.8)	7.1 (5.6)	14.3 (8.3) ^b	20.9 (6.9) ^b	18.4 (5.8)
SEV	4.0 (4.5) ^a	2.5 (3.5)	9.5 (8.6) ^b	16.5 (8.4) ^b	13.3 (6.5)
RO	2.1 (2.3)	1.6 (2.0)	5.1 (3.4)	6.1 (3.9)	5.9 (2.7)

^a Significant difference from honest clinical patients reported in Rogers et al. (1992) ($p < .05$)

^b Significant difference from malingering/simulating sample reported in Rogers et al. (1992) ($p < .05$)

endorsed 20 or more of the items. Most of the community participants ($n = 21$; 72.4%) had at least one scale in the "definite" range and four (13.8%) more participants obtained three or more scales in the "probable" malingering range. Thus, 25 of the 29 community participants (86.2%) in the "malingering" group were correctly identified by the SIRS-A.

Accuracy Predicting Response Style

Using the criteria recommended by Rogers et al. (1992) for classifying malingering (i.e., either one scale in the "definite" malingering range or three scales in the "probable" range), sensitivity was 86.2% and specificity was 85.1% with an overall accuracy rate of 85.3%. Unfortunately, the SIRS-A yielded a relatively low PPA rate (65.8%). The chi-square test of association was used to compare the predictive accuracy obtained in this study with the original SIRS validation studies. Expected frequency estimates were based on data presented in the SIRS manual (i.e., sensitivity of 48.5% and specificity of 99.5%; calculated from classification accuracies reported on page 24). This analysis revealed a significantly different pattern of classification results for the SIRS-A, $\chi^2(1, N = 116) = 376.84$, $p < .0001$, $\phi = 1.80$, which resulted from both higher than expected false positive and sensitivity rates. False positive predictions were made for 13 of 87 mentally ill participants whereas the SIRS normative data would predict that less than one such errors would be made in a sample of this size. On the other hand, accurate identification of malingerers (i.e., sensitivity) occurred in 25 of 29 cases, whereas the expected rate was only 14 of 29.

Receiver Operating Characteristics of the SIRS-A

Receiver operating characteristics (ROC) analysis was used to evaluate the classification accuracy of the SIRS-A based on the total score (i.e., sum of all eight scales) and each of the individual subscales. The SIRS-A total score performed significantly better than chance, with an Area Under the Curve (AUC) of .92 (95% CI: .85–.99). A cut-off score of 48 yielded sensitivity of 93.1% and specificity of 77.0%, with an overall accuracy of 81.0%. Raising the cut-off score to 70 increased specificity to 88.5% but decreased sensitivity to 86.2%, with an overall accuracy of 87.9%.

ROC curves also indicated that each of the eight individual SIRS-A scales performed significantly better than chance in predicting malingering, with a mean AUC of .88 (range: .79–.92; $p < .001$ for each analysis). The AUC data for each of the scales and total score are presented in Table 3, along with the results of comparable analyses using the psychiatric malingering sample.

Comparing “Honest” Psychiatric Patients to “Malingering” Psychiatric Participants

Psychiatric patients instructed to mangle obtained a total score on the SIRS-A of 67.67 (SD = 46.6, range: 7–172), more than twice as high as the “honest” psychiatric sample (mean = 30.16), $t(109) = 3.80$, $p < .001$, $d = 0.73$. However, the average SIRS-A total score for the psychiatric patient malingering sample was 55.3 points lower than for the community malingering sample, $t(51) = 2.83$, $p = .006$, $d = .79$. In fact, only five psychiatric patient malingerers (20.8%) obtained SIRS-A total scores above the mean of the community malingering sample (102 or greater) and 9 (37.5%) fell above the cut-off of 70

identified as the optimal cut-off based on the analyses above. Although the psychiatric malingering sample did not attain elevated total scores of the same magnitude found in the community sample, most ($n = 20$; 83.3%) still endorsed more than 20 items. Seven (29.2%) of the psychiatric patients instructed to mangle scored within the “definite” malingering range on one or more of the SIRS-A scales and an additional five (20.8%) patients obtained three or more scales in the “probable” malingering range. Therefore, half ($n = 12$; 50%) of the participants instructed to mangle were not identified as malingering by the SIRS-A. This rate of “false negatives” was significantly greater than for the community malingering sample (4 of 29, 13.8%), $\chi^2 = 8.17$, $p = .005$, $\phi = .39$. This difference between psychiatric and community malingerers is further highlighted by contrasting the mean scores on each of the eight scales, and the proportion of respondents for each scale whose scores were elevated to the probable or definite ranges (see Table 4). The mean scale score was significantly lower on five of the eight scales for the psychiatric malingerers than those in the community instructed to mangle mental illness. Although the psychiatric malingering sample typically endorsed fewer items than those in the community sample, the mean score attained on all eight scales was significantly higher for the psychiatric malingerers than those patients instructed to respond honestly.

Accuracy at Predicting Response Style

Using the original SIRS scoring criteria described above, sensitivity was 50.0% and specificity was 85.1%, with an overall accuracy rate of 77.5%. This accuracy rate corresponds to a PPA of 48% and a NPA of 86%. These accuracy rates were significantly poorer than the original

Table 3 Area under the curve for the eight SIRS-A scales and total score for the community and psychiatric malingerers

Scale	Community malingerers		Psychiatric malingerers	
	AUC	95% CI	AUC	95% CI
SC	.92	.87–.97	.80	.69–.90
BL	.92	.85–.99	.77	.66–.88
IA	.92	.87–.96	.73	.61–.86
SEL	.91	.83–.98	.71	.59–.83
SEV	.89	.80–.98	.69	.55–.83
RS	.88	.81–.95	.72	.60–.84
SU	.84	.75–.93	.67	.54–.80
RO	.79	.68–.91	.77	.65–.89
Total score	.92	.85–.99	.76	.65–.87

The AUC of all scales and total score were significantly better than chance, contrasting malingerers in each group to the honest psychiatric sample ($p < .05$)

Table 4 Sample comparisons on scale scores and elevations

Scale	Honest psychiatric		Malingering psychiatric		Malingering community	
	Mean	% Prob/Def	Mean	% Prob/Def	Mean	% Prob/Def
RS	2.0 ^a	12.6/2.4	5.0 ^b	16.7/25.0	8.0 ^c	27.6/48.3
SC	1.6 ^a	5.7/1.1	6.0 ^b	8.3/20.8	8.4 ^b	55.2/20.7
IA	.9 ^a	1.1/1.1	4.0 ^b	12.5/20.8	5.3 ^b	10.3/31.0
BL	3.9 ^a	11.5/0.0	11.1 ^b	25.0/12.5	17.9 ^c	69.0/20.7
SU	8.0 ^a	12.6/0.0	12.7 ^b	29.2/8.3	19.0 ^c	37.9/24.1
SEL	8.3 ^a	9.2/0.0	14.3 ^b	25.0/4.2	20.9 ^c	72.4/0.0
SEV	4.0 ^a	12.6/0.0	9.5 ^b	29.2/16.7	16.5 ^c	31.0/51.7
RO	2.1 ^a	6.9/ 0.0	5.1 ^b	37.5/0	6.1 ^b	44.8/6.9
Total score	30.2 ^a	N/A	67.7 ^b	N/A	102.0 ^c	N/A

Note: Different subscripts represent significant mean differences between groups ($p < .05$)

SIRS validation data, $\chi^2(1, N = 111) = 398.75, p < .0001$, $\phi = 1.90$ which is reflective of the elevated rate of false positives in the present study (noted above). The sensitivity rate in this sample was equivalent to the expected frequency based on the SIRS normative data (51%).

Receiver Operating Characteristics of the SIRS-A

Receiver operating characteristics (ROC) analyses indicated that the SIRS-A total score performed significantly better than chance at predicting malingering, with an Area Under the Curve (AUC) of .76 (95% CI: .65–.87). As in the previous analysis of community malingerers, the optimal cut-off score was 48. Applying this cut-off with psychiatric patients as the malingering comparison yielded a sensitivity of 70.8% and a specificity of 77.0%, with an overall accuracy of 75.7%. Raising the cut-off score to 60 increased specificity to 82.8% but decreased sensitivity to 54.2%.

ROC curves were analyzed for each of the eight individual SIRS-A scales. Each of the scales performed significantly better than chance ($p < .05$), with a mean AUC of .73 (range: .67–.80). As displayed in Table 3, AUCs were uniformly higher when assessed by comparing honest patients to malingering community participants, than when the latter was replaced with malingering psychiatric patients.

Discussion

These studies provide preliminary support for the use of an abbreviated version the SIRS (the SIRS-A) for detecting malingering. Unlike prior studies that have analyzed abbreviated versions of the SIRS, the SIRS-A includes all of the primary scale questions and therefore can be scored and interpreted using the same guidelines designed for the original SIRS and directly compared against the criterion. Additionally, the present study used a simulation design with two different samples of individuals instructed to malingering (psychiatric patients and a community sample), and a realistic base rate of malingering. Although the psychometric properties (e.g., scale internal consistency) of the SIRS-A appeared comparable to the full-length SIRS, a comparison of individual scale scores between our samples and Rogers' original validation studies indicate important differences in response patterns. The data described here demonstrated equivalent or even improved sensitivity rates when compared to the original SIRS validation data (Rogers et al. 1992), but we found substantially higher rates of misclassification (i.e., lower specificity and positive predictive accuracy). Specifically, 10–15% of chronically ill psychiatric patients who were asked to respond

honestly to the test materials were classified by the SIRS-A as malingering.

The base rates of malingering in this study were intentionally established to reflect the probability of occurrence typical in forensic settings (e.g., 17% reported by Rogers et al. 1998). This balance results in realistic predictive accuracy rates that forensic evaluators could expect in clinical settings. However, comparisons to published data also emphasized sensitivity and specificity rates, as these indices are unaffected by base rate (as opposed to positive and negative predictive accuracies). Sensitivity rates were most impressive in the study of community participants instructed to malingering mental illness. In fact, sensitivity in detecting malingering among this sample of community participants was superior to the sensitivity rates described by Rogers et al. (1992) in their published normative data (86% vs. 49%, respectively). Assuming that this community sample (72% of whom had prior arrests) is roughly comparable to persons who might be charged with a criminal offense, these findings support the utility of the SIRS-A for detecting malingering in forensic settings.

Sensitivity rates for detecting malingering among the psychiatric patients was substantially lower than for the community sample, although still comparable to Rogers' normative data (50% vs. 49%). One possibility for the decreased sensitivity among genuine psychiatric patients asked to malingering is that bona fide patients exaggerating psychopathology are better adept at alluding detection than non-mentally ill individuals. Community members appeared more likely to endorse symptoms indiscriminately, endorsing almost twice as many SIRS-A items as the psychiatric patients instructed to malingering. However, it is also possible that some proportion of the mentally ill individuals who were not detected as malingerers may have been too impaired to malingering consistently. Yet the finding that very few of the psychiatric patients endorsed large numbers of SIRS-A items supports the former hypothesis, as even a large proportion of individuals who did not actually "malingering" would not account for the large difference in the number of SIRS-A items endorsed between groups. Thus, even significantly impaired psychiatric patients appear capable of malingering more successfully on the SIRS-A than a "typical" community sample representing a range of demographic characteristics, mental health, and incarceration histories.

Perhaps more importantly, specificity and PPA rates observed in this sample of genuinely impaired psychiatric patients were significantly worse than anticipated based on published normative data for the SIRS. Whereas Rogers and colleagues' report near-perfect specificity based on their criterion for classifying malingering (i.e., at least one scale in the "definite" malingering range or three in the "probable" range), specificity in this sample was only

85%. Several possibilities might explain why the SIRS-A failed to achieve expected specificity rates, including the use of an abbreviated version of the SIRS. Norris and May (1998) found that accuracy diminished when they administered a paper and pencil screening version of the SIRS compared to when they administered the full-length SIRS and extracted the responses to the screening items, suggesting that inclusion of distracter items may help to reduce misclassifications. Alternatively, the SIRS-A (and possibly the SIRS) may perform worse when used with severely mentally ill patients rather than less impaired forensic or community samples (which forms the basis for much of the published normative data). Although relatively few genuinely impaired psychiatric patients obtained subscale elevations in “definite” range, a substantial number obtained scale scores in the “probable” range. Further research with severely ill psychiatric patients is necessary to determine if this elevated risk of false classification extends to the SIRS, or is limited to the SIRS-A.

It is also possible that some of the participants who were asked to respond honestly chose to instead exaggerate their symptoms. Indeed, past research has suggested that roughly 7% of psychiatric patients in general clinical settings may be malingering mental illness (Rogers 1997). If a few such individuals were included in the present study, they would be indistinguishable from the malingering group, and hence would have been *correctly* classified as malingering. This explanation, although possible, appears unlikely since virtually all of the patients studied had been repeatedly hospitalized and were receiving outpatient mental health services, often for many years. Moreover, all of the participants were explicitly informed that the results of the tests were confidential and would not be shared with clinical staff, giving them no motivation to exaggerate their symptoms. Finally, the proportion of false positives (34–52%) far exceeds estimates of malingering in general clinical settings, suggesting that this explanation is unlikely to account for the inflated false positive rate.

Finally, the finding that the SIRS-A performed worse than previously published studies using the SIRS may reflect the nature of the studies to which it was compared. The SIRS-A was contrasted to the results data described in the SIRS manual. However, because the original scoring criteria were developed from these very studies, there is a high likelihood that the cut-off scores over-fit the data and subsequent studies would inevitably fail to duplicate these impressive accuracy rates. In addition, the SIRS normative data merged samples of psychiatric patients, criminal offenders, and community subjects, likely resulting in a higher accuracy rate for the SIRS than might be found with only psychiatric patients. Thus, the performance of the SIRS-A might have appeared more favorable if contrasted to other studies not used in the initial validation of the

SIRS, or with only subsamples of the normative group such as psychiatric patients. However, studies conducted subsequent to the publication of the SIRS manual have been limited to forensic samples (e.g., Gothard et al. 1995). Therefore, only the original SIRS studies were deemed comparable to the present analyses.

Interestingly, classification accuracy based on the total number of items endorsed was roughly equivalent to the criterion established by Rogers et al. (1992) using scale elevations in the definite or probable range. Sensitivity using the total score was also comparable or superior in the present analyses compared to data described by Rogers and colleagues. We found sensitivity rates of 86% in the community sample and 54% in the psychiatric sample compared to 56% in Rogers’ analyses (Rogers et al. 1992), although these analyses all used different cut-off scores (in part because Rogers et al. utilized the entire SIRS measure). Thus, it appears that the total score is potentially a more accurate and more useful index of malingering in the SIRS-A compared to the scale classifications.

The current study has several limitations. First, there are problems inherent in any study that employs a simulation methodology rather than utilizing known criterion groups. Rogers and Cruise (1998) coined the phrase the “simulation-malingering paradox” to refer to the questionable generalizability resulting from studies that direct persons to comply with instructions to feign mental illness in order to understand persons who malingering when instructed to comply. For example, the issue of motivation in actual testing situations may have an impact on how malingerers feign their symptoms. In the real world, malingerers stand to gain significant benefits, such as financial compensation or avoiding incarceration or trial, if they successfully feign illness. Without these tangible benefits, the participants in this study may have responded differently than psychiatric patients or correctional examinees who try to exaggerate symptoms for external benefit. Other studies have attempted to offset some of these concerns by offering financial incentives to participants who malingering convincingly. If the participants in this study had received external motivation to exaggerate symptoms, it is possible that the SIRS-A would have had greater difficulty accurately identifying them. However, whether these incentives, typically in the form of a bonus of some type, result in performance truly akin to persons malingering for the purpose of evading trial or prison has been questioned and may be less influential than often assumed (Miller 1999).

The results of the SIRS-A must also be tempered by the fact that the SIRS was not administered concurrently. Without the performance of the SIRS as an immediate comparison, it is not possible to determine whether the results obtained in this study, particularly the high per-

centage of false positives, are due to the sample tested, the omission of many original SIRS items, or other study design issues. Future studies would benefit from incorporating both measures to gain greater clarity on this issue, as well as provide a direct comparison of the accuracy of the SIRS-A to its predecessor.

Finally, although the present study excluded participants who clearly did not understand the instructions, some participants may have still had difficulty understanding and remembering the instructions to malingering. Thus, as noted above, some proportion of the “successful” malingerers may have actually responded honestly (i.e., were accurately classified as honest). Conversely, as noted above, some genuine psychiatric patients may have been malingering their symptoms despite being asked to respond honestly. While this risk appears inevitable when asking psychiatric patients to malingering mental illness, these factors are likely to have only influenced a handful of subjects and do not invalidate the broader pattern of findings. Nevertheless, future malingering studies using genuine psychiatric patients may benefit from incorporating more extensive manipulation checks to ensure comprehension and compliance. Additionally, research using samples of persons independently suspected of malingering (i.e., a known-groups design) would also help clarify these issues.

Malingering is relatively common in both clinical and forensic settings and the SIRS-A may be one means to help practitioners and evaluators rapidly and in a cost-effective manner determine what services are best suited for patients, offenders, and examinees. Overall, these results are promising and indicate that the SIRS-A may serve as an effective and rapid means of assessing malingering without requiring extensive time in settings where resources are often limited. Despite this promise, future research is necessary to determine whether the high false positive rate in this study resulted from eliminating the extraneous SIRS items or because of the severity of mental illness in this study’s sample. Particularly given the potential significance of a “false positive” assessment (i.e., concluding that a genuinely ill individual is malingering), the accuracy of malingering instruments with this population requires further study.

Appendix Scales on the SIRS-A

Scales	Description
Rare symptoms (RS)	Symptoms that rarely are found in actual psychiatric populations
Symptom combinations (SC)	Grouping of two common symptoms that rarely occur simultaneously
Improbable and absurd Symptoms (IA)	Symptoms that are so outrageous as to be highly implausible

Appendix continued

Scales	Description
Blatant symptoms (BL)	Symptoms that are such obvious signs of mental illness as to be easily recognizable to an untrained person
Subtle symptoms (SU)	Symptoms that do not readily indicate mental illness to an untrained person, but may instead be presumed to be everyday problems
Selectivity of symptoms (SEL)	Combines the blatant and subtle symptoms and assesses the extent to which a responder will endorse indiscriminant items
Severity of symptoms (SEV)	Measure of the number of blatant and subtle items reported to be unbearable or extremely painful
Reported versus observed symptoms (R/O)	Comparison between symptoms endorsed and actual observable behavior

References

- Arbisi, P. A., & Ben-Porath, Y. S. (1998). The ability of the Minnesota Multiphasic Personality Inventory-2 validity scales to detect fake-bad responses in psychiatric inpatients. *Psychological Assessment*, 10(3), 221–228.
- 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, 1478–1481.
- Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *MMPI-2: Manual for administration and scoring*. Minneapolis: University of Minnesota Press.
- Gothard, S., Viglione, D. J., Meloy, J. R., & Sherman, M. (1995). Detection of malingering in competency to stand trial evaluations. *Law and Human Behavior*, 19(5), 493–505.
- Gough, H. G. (1950). The F minus K dissimulation index for the Minnesota Multiphasic Personality Inventory. *Journal of Consulting Psychology*, 14, 408–413.
- Greene, R. L. (1997). Assessment of malingering and defensiveness by multiscale inventories. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (2nd ed., pp. 169–207). New York, NY: Guilford Press.
- Hathaway, S. R., & McKinley, J. C. (1940). A multiphasic personality schedule (Minnesota): I. Construction of the schedule. *Journal of Psychology*, 10, 249–254.
- Lally, S. J. (2003). What tests are acceptable for use in forensic evaluations? A survey of experts. *Professional Psychology: Research and Practice*, 34(5), 491–498.
- Meyer, R. G., & Deitsch, S. M. (1995). The assessment of malingering in psychodiagnostic evaluations: Research-based concepts and methods for consultants. *Consulting Psychology Journal*, 47(4), 234–245.
- Miller, H. A. (1999). *The development of the Miller’s forensic assessment of symptoms test: A measure of malingering mental illness*. Florida State University College of Arts and Sciences.
- Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, 24(8), 1094–1102.
- Norris, M. P., & May, M. C. (1998). Screening for malingering in a correctional setting. *Law and Human Behavior*, 22(3), 315–323.

- Overall, J. E., & Gomez-Mont, F. (1974). The MMPI-168 for psychiatric screening. *Educational and Psychological Measurement*, 34, 315–319.
- Ratner, I. K. H. (1984). The efficacy and accuracy of the MMPI-168 in a forensic setting. Tennessee State University.
- Rogers, R. (1997). Structured interviews and dissimulation. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (2nd ed., pp. 301–327). New York, NY: Guilford Press.
- Rogers, R., Bagby, R. M., & Dickens, S. E. (1992). *SIRS: Structured interview of reported symptoms: Professional manual*. Odessa, FL: Psychological Assessment Resources, Inc.
- Rogers, R., Gillis, J. R., & Bagby, R. M. (1990). The SIRS as a measure of malingering: A validation study with a correctional sample. *Behavioral Sciences and the Law*, 8, 85–92.
- Rogers, R., Gillis, J. R., Bagby, R. M., & Monteiro, E. (1991). Detection of malingering on the structured interview of reported symptoms (SIRS): A study of coached and uncoached simulators. *Psychological Assessment*, 3(4), 673–677.
- Rogers, R., Gillis, J. R., Dickens, S. E., & Bagby, R. M. (1991). Standardized assessment of malingering: Validation of the structured interview of reported symptoms. *Psychological Assessment*, 3(1), 89–96.
- Rogers, R., Hinds, J. D., & Sewell, K. W. (1996). Feigning psychopathology among adolescent offenders: Validation of the SIRS, MMPI-A, and SIMS. *Journal of Personality Assessment*, 67(2), 244–257.
- Rogers, R., Salekin, R. T., Sewell, K. W., Goldstein, A., & Leonard, K. (1998). A comparison of forensic and nonforensic malingerers: A prototypical analysis of explanatory models. *Law and Human Behavior*, 22(4), 353–367.
- Rosenfeld, B., Sands, S. A., & Van Gorp, W. G. (2000). Have we forgotten the base rate problem? Methodological issues in the detection of distortion. *Archives of Clinical Neuropsychology*, 15(4), 349–359.
- Story, D. L. (2000). *Validation of a short form of the structured interview of reported symptoms (SIRS)*. Miami, Florida: Carlos Albizu University.
- Walters, G. D., White, T. W., & Greene, R. L. (1988). Use of the MMPI to identify malingering and exaggeration of psychiatric symptomatology in male prison inmates. *Journal of Consulting and Clinical Psychology*, 56(1), 111–117.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.