Archives of Clinical Neuropsychology 21 (2006) 651–656

Improving the clinical utility of the SIRS cognitive items: Preliminary reliability, validity, and normative data in pretrial and clinical samples

Timothy F. Wynkoop [a,*], Richard I. Frederick [b], Michelle Hoy [b]

[a] Court Diagnostic & Treatment Center, Toledo and Bowling Green State University, OH, United States
[b] U.S. Medical Center for Federal Prisoners, Springfield, MO, United States

Accepted 27 June 2006

Abstract

The utility of the Structured Interview of Reported Symptoms [SIRS; Rogers, R., Bagby, R. M., & Dickens, S. E. (1992). *Structured Interview of Reported Symptoms professional manual*. Odessa, FL: Psychological Assessment Resources], Improbable Failure Rate (IF) scale in pretrial ($N = 64$) and clinical ($N = 153$) samples was explored. Internal consistencies of the IF items were $\alpha = .81$ and $\alpha = .92$, respectively, with split-half reliabilities of .89 and .84, respectively. The IF scale loaded distinctly from the SIRS primary (psychosis) scales in principal components analysis, and the pattern of performances among clinical groups were as expected (e.g., normal controls outperformed patients with focal lesions and dementia, CHI patients outperformed dementia patients). Performance on the IF items was correlated with MMSE score ($r = -.59$, $p \leq .001$), age ($r = .53$, $p \leq .001$), and education ($r = .31$, $p \leq .001$), but not with Barona estimated FSIQ ($r = -.14$, $p = .08$). For normals and psychiatric patients there was a clear falling off of IF error scores after 1 error, and for CHI patients the fall off began after 2 errors. Results lend credence to the SIRS IF scale as an acceptable screening measure for feigned cognitive abilities.

Keywords: Malingering; Cognition; Cognitive deficit; SIRS

Malingering is the intentional fabrication or gross exaggeration of symptoms in an effort to gain some goal or to avoid some punishment (American Psychiatric Association [APA], 2000). It is a classification consideration in many instances of clinical and forensic neuropsychological assessments (Bush et al., 2005; Hom & Denney, 2002; cf. Rogers, 1997a). Survey results suggest that 29% of personal injury, 30% of disability, 8% of medical, and 19% of criminal evaluation cases may be suspicious for feigning or exaggeration (Mittenberg, Patton, Canyock, & Condit, 2002). One study of civil litigants found that 41% had probably feigned or exaggerated cognitive deficits for financial gain (Greiffenstein, Baker, & Gola, 1994).

The variability in rates of feigned psychiatric or cognitive problems are caused by the pitfalls of epidemiological research in general, such as varying settings and referral sources, differing decision making thresholds, and incomplete access to collateral information (cf. Wynkoop, Capps, & Priest, 1995). Additionally, different persons deceive differently (Pankratz, 1998), and it seems probable that a malingerer's style of feigning or exaggerating can change from situation

* Correspondence to: P.O. Box 332, Maumee, OH 43537, United States.
E-mail address: tfwynkoop@yahoo.com (T.F. Wynkoop).

0887-6177/$ – see front matter
doi:10.1016/j.acn.2006.06.009

to situation or from test to test (Bush et al., 2005; Wynkoop & Denney, 1999). Some malingerers are very sophisticated, and perhaps even coached, while others are easily exposed (Tan, Slick, Strauss, & Hultsch, 2002). Some are selective in the symptoms that they wish to exaggerate or feign, while others take a broader more varied approach hoping that one of their problems will provide them some benefit or absolve them from some degree of responsibility. Consequently, there is a need for instruments designed to detect dissimulation that are broad in their approach.

1. Structured Interview of Reported Symptoms (SIRS)

The SIRS (Rogers, Bagby, & Dickens, 1992) is a 172-item structured interview designed to detect and to characterize feigned mental disorders. It has eight primary scales: Rare Symptoms, Symptom Combinations, Improbable and Absurd Symptoms, Blatant Symptoms, Subtle Symptoms, Selectivity of Symptoms, Severity of Symptoms, and Reported versus Observed Symptoms. Selectivity of Symptoms and Severity of Symptoms overlap with items from other scales, but only in specificity and intensity of symptoms, not in item content. There are also five supplementary scales: Direct Appraisal of Honesty, Defensive Symptoms, Overly Specified Symptoms, Symptom Onset and Resolution, and Inconsistency of Symptoms. The primary scales reliably differentiate feigners and honest responders, while the supplementary scales help to characterize response styles. The SIRS authors report a correct classification rate of 94.3% in a clinical sample with a base rate of 25% for suspected malingering of psychosis, and correct classification rates of 89.8% in a calibration sample with a base rate of 64% for simulated malingered psychosis and 88.3% in a cross validation sample with a base rate of 70% for simulated malingered psychosis.

Although not a test of feigned cognitive deficit per se, the SIRS contains four groups of five items each that measure generation of opposites (numbers 57 and 143) and rhymes (numbers 58 and 144; each with five items, a–e) known as the Improbable Failure Rate scale (IF; Rogers et al., 1992). The items are scored as errors, using a floor effect detection strategy (cf. Rogers, 1997b). Higher IF scores suggest feigning of cognitive deficit. However, the reliability and validity of the IF scale has yet to be reported in the literature, despite remaining in the SIRS administration booklet. In the analyses that follow, we present preliminary normative data for the SIRS IF scale in forensic and clinical samples, address the reliability and validity of the IF scale in these groups, and make suggestions for its use.

2. Method

2.1. Participants

The pretrial sample consisted of 64 men referred from across the United States to a locked forensic unit for pretrial evaluation. Their average age was 34.9 (S.D. = 9.4) and their average education was 10.6 (S.D. = 2.5) years. The entire SIRS was administered to this sample. Thirty-six (56.3%) of the men had three or more SIRS scale scores falling in the probable feigning range, meeting SIRS classification criteria for "definite feigning" (i.e., with 98.9% certainty; Rogers et al., 1992, p. 24).

The clinical sample consisted of 153 inpatients and outpatients (see Table 1 for demographics). Non-neurologically involved (i.e., "normal") participants were those without central nervous system involvement (e.g., peripheral demyelinating diseases, non-psychotic behavioral problems). Closed head injury (CHI) patients had suffered substantial posttraumatic amnesia and/or had evidence of traumatic brain injury on neuroimaging, without penetrating injury or skull fracture. Those with focal lesions included occlusive CVA ($N = 18$) or tumor ($N = 4$). Those with encephalopathy suffered infections ($N = 16$) or spontaneous bleeds ($N = 7$; seen in a subacute rehabilitation setting where general

Table 1
Clinical sample demographics ($N = 153$)

Variable	Mean	S.D.	Minimum	Maximum
Age	54.3	21.6	18	91
Education	11.9	2.3	4	18
Barona FSIQ	101.6	7.8	82.2	120.9
MMSE	25.9	4.5	9	30
IF (errors)	4.2	5.0	0	17

Table 2
Clinical group means ($N = 153$)

Group (N)	Age	Education	Barona FSIQ	MMSE	IF errors
Normal (12)	43.7 (17.0)	12.2 (.9)	101.2 (5.1)	28.5 (.9)	.5 (.7)
Closed head injury (28)	34.9 (16.0)	12.3 (2.7)	101.6 (8.3)	27.8 (2.5)	1.5 (2.5)
Focal lesions (22)	67.5 (17.8)	12.6 (2.2)	105.3 (7.8)	23.3 (4.7)	4.8 (4.5)
Encephalopathy (23)	56.9 (18.0)	11.7 (2.3)	100.8 (8.9)	23.4 (5.9)	4.4 (5.3)
Dementia (39)	73.8 (10.8)	11.2 (2.6)	101.6 (7.2)	24.4 (4.6)	6.9 (5.2)
Psychiatric (29)	38.9 (13.6)	11.9 (1.9)	99.6 (7.6)	28.9 (1.4)	1.0 (1.6)

Standard deviations in parentheses.

encephalopathy was still the primary clinical presentation). Dementia patients suffered cortical and/or progressive cerebrovascular decline ($N = 30$), demyelinating disease ($N = 5$), subcortical gray nuclei disease ($N = 3$), and normal pressure hydrocephalus ($N = 1$), all with diagnoses confirmed via neuroimaging, medical consults, and/or additional neuropsychological data. Psychiatric patients suffered a variety of mood disorders ($N = 23$) and/or psychoses ($N = 6$), with psychotic symptoms being in at least fair remission (i.e., none were floridly psychotic at the time of testing), with diagnoses arising from clinical interview, psychiatric records, and/or personality testing (i.e., MMPI, PAI). Patients did not demonstrate obvious problems with comprehension or expression, and none had obvious motivation to malinger. Only the IF items from the SIRS were administered to the clinical samples, in the same staggered fashion as found in the SIRS administration booklet.

Normals were younger than those with focal lesions or dementing conditions; CHI and psychiatric patients were younger than those with focal lesions, encephalopathy, and dementia; and encephalopathic patients were younger than dementia patients ($F[5,147] = 33.1$, $p \leq .001$; Table 2 for means and standard deviations). Patients with focal lesions, encephalopathy, and dementias scored substantially lower than normals, CHI, and psychiatric patients on the Mini-Mental State Exam (MMSE; Folstein, Folstein, & McHugh, 1975; $F[5,147] = 11.0$, $p \leq .001$, Bonferroni correction). The groups did not differ significantly with regard to education ($F[5,147] = 1.3$, $p = .274$) or Barona FSIQ (Barona, Reynolds, & Chastain, 1984; $F[5,147] = 1.5$, $p = .204$).

3. Statistical analyses

3.1. Reliability

Internal consistency of the IF in the pretrial sample ($N = 64$) was $\alpha = .81$, with split-half reliabilities of .89 for each half. All four of the IF item groups contributed to the internal consistency of the IF scale, although opposites were slightly stronger in this regard (i.e., their deletion would decrease α).

Internal consistency of the IF item blocks in clinical samples ($N = 153$) was $\alpha = .84$. As with the forensic evaluation participants, all four of the IF item groups contributed to the internal consistency of the IF scale (α ranged from .72 to .84), although opposites were slightly stronger in this regard. Internal consistency using individual IF items was $\alpha = .92$. Deletion of any of the individual items did not change internal consistency appreciably (α ranged from .91 to .92). Split-half reliabilities were .83 and .85.

3.2. Construct validity

The IF scale was developed on a conceptually sound theoretical basis. Its basis was response bias detection theory (Rogers et al., 1992) as an extension of neuropsychological assessment principles. Consequently, the IF items had the benefit of well-grounded theory and expert construction from their inception. The inclusion of the cognitive items in the SIRS likely reflects the willingness of the authors to explore a large variety of detection strategies (e.g., endorsing rare symptoms, endorsing improbable and absurd symptoms, floor effect; Rogers, Harrell, & Liff, 1993).

3.3. Pretrial sample

The eight SIRS primary scales and the IF scale were subjected to principal components analysis (PCA) using the forensic sample. Because we were interested in characterizing the constructs underlying the classification of feigning

Table 3
Principal components analysis

SIRS scales	Components	
	1	2
Rare Symptoms	.873	.247
Symptom Combinations	.724	.210
Improbable and Absurd	.682	.436
Blatant Symptoms	.924	−.116
Subtle Symptoms	.821	−.337
Selectivity of Symptoms	.926	−.227
Severity of Symptoms	.926	−.227
Reported vs. Observed	.694	−.110
Improbable Failure Rate	.248	.836

(as opposed to characterizing feigning itself), and to maximize the subject to variable ratio, it made sense to use only the SIRS primary scales.

Two components emerged (eigenvalues ≥ 1), accounting for 75.2% of the variance (Table 3). The first component, accounting for 61.5% of the variance, comprised all eight of the SIRS primary scales which measure malingered psychiatric symptoms (all loadings ≥ .68). The IF scale was the only very strong loading on the second component ($r = .84$). Interestingly, Improbable and Absurd Symptoms also loaded secondarily on the second component ($r = .44$), sharing ample variance with the IF scale. However, this loading diminished ($r = .30$) when we conducted a second PCA, separating the IF into opposites ($r = .81$) and rhymes ($r = .81$), with all other component loadings remaining essentially unchanged.

IF scores ranged from 0 to 17 (errors), with an average of 4.8 (S.D. = 4.7) errors. Participants whose IF score fell in the lower quartile had an average of 0 errors. Half of the participants had four or fewer (≤4) errors. The upper quartile error range was 7–17. For those whose SIRS primary scales did not suggest probable feigning or exaggeration ($N = 26$), IF scores ranged from 0 to 17, with an average of 4.2 (S.D. = 5.4) errors. Half of these participants made one or less (≤1) errors (mode of zero errors). Interestingly, the correlation between IF and number of SIRS primary scale scores falling in the probable feigning range was low ($r = .18$, $p = .167$), despite the high likelihood of feigning or exaggeration of psychiatric symptoms in this group, suggesting that these participants tended not to feign psychosis and cognitive deficits indiscriminately.

Zero order correlations between IF score and other SIRS primary scales were significant for Rare Symptoms ($r = .35$, $p = .005$), and for Improbable and Absurd Symptoms ($r = .32$, $p = .010$). There were also significant relationships with the supplementary scales of Defensive Symptoms ($r = -.26$, $p = .041$) and Inconsistency of Symptoms ($r = .52$, $p < .001$). These correlations could be indicative of confusion, or of a feigning strategy to appear confused. There was not a significant relationship between the IF score and education ($r = .17$, $p = .192$), nor with SIRS completion time ($r = -.13$, $p = .341$).

3.4. Clinical samples

Psychiatric patients outperformed dementia, focal lesion, and encephalopathic patients on the IF items, normals outperformed patients with focal lesions and dementia, and CHI patients outperformed dementia patients. Zero order correlations between IF errors and MMSE score ($r = -.59$, $p \leq .001$), age ($r = .53$, $p \leq .001$), and education ($r = .31$, $p \leq .001$) were significant, but not between IF errors and Barona estimated FSIQ ($r = -.14$, $p = .08$; order of entry confirmed on stepwise linear regression). These data suggest that IF performance in clinical populations may be dependent on current cognitive functioning and age, not necessarily on premorbid psychometric intelligence (i.e., in our sample—keeping in mind the fairly large Barona error variance; and our own clinical observations have been that persons with defective intelligence have legitimate difficulty with the IF items). The relationship between IF errors and age, and MMSE scores, remained even when MMSE scores and age were covaried (ANCOVA). These findings were not unexpected and suggest that aging exerts influence on ability to generate opposites and/or rhymes, perhaps as a function of slowed mentation.

Review of raw data and cumulative percentages suggested that for normals and psychiatric patients there was a clear falling off of IF error scores after 1 error, and for CHI patients the fall off began after 2 errors. For example, 11 of the 12 (92%) normals made 0 or 1 errors (7, or 58%, made no errors); 24 of the 29 (83%) psychiatric patients made 0 or 1 errors (15, or 52% made no errors); and 20 of the 28 (71%) CHI patients made 0 or 1 errors, with 24 (86%) making 0–2 errors (11, or 39% made no errors). Whereas, only 36% of focal lesion patients, 52% of encephalopathic patients, and 18% of dementia patients made only 0 or 1 errors. This pattern was not unexpected on the basis of general cognitive functioning as measured by the MMSE.

Interestingly, no one from the clinical sample missed item 143 (a), an opposites item, even with MMSE scores as low as 9. Other rarely missed items were 57 (c), (d), and (e), all opposites.

4. Discussion

The IF scale demonstrated adequate reliability in our pretrial and patient samples, and may be useful in helping evaluators to determine whether they should further evaluate cognitive capacities, including feigned cognitive deficits. The IF scale appears to measure fabrication and/or exaggeration of cognitive problems distinct from fabrication and/or exaggeration of psychotic symptoms. Exaggeration or fabrication of cognitive problems of the type measured by the IF seems to be related to a malingering strategy of appearing confused, endorsing rarely endorsed and/or absurd symptoms, and of eagerness to offer these symptoms to the examiner.

The IF scale retained its floor effect appeal in our normals and psychiatric patients (the populations for which the SIRS was intended), but not so in our patients with focal lesions, encephalopathy, or dementia. However, this was not unexpected. Conceptually, the relationship between IF errors and MMSE scores, and how this relationship expresses itself across various patient groups, lends additional credence to the construct validity of the IF scale. Pragmatically, for clinical patients who may be exaggerating cognitive problems the underlying detection strategy simply shifts from *floor effect* to *typical performance* for the particular patient groups. While the typical performance detection strategy may seem less powerful than the floor effect strategy in terms of certainty, or evidence of intentionally suppressed cognitive performance (cf. Wynkoop & Denney, 1999), the strong correlation between MMSE and IF scores also allows for analysis of *consistency of performance* (cf. Rogers, 1997b) as an additional method of detecting intentionally suppressed performance using our data.

If an evaluator uses these results to interpret the IF scale when administering the SIRS, we suggest also administering the MMSE and calculating a Barona IQ estimate so that more direct comparisons using our data can be made. Take advantage of IF *floor effect*, if appropriate, and/or see if the data can speak to *typical* (or *atypical*) *performance* or to *inconsistent performance* across settings. For example, one way to use the IF scale is to examine whether the person is suppressing his/her MMSE and IF scores to the point of being unbelievable in relation to his/her daily living activities.

As these analyses demonstrate, persons without neuropsychological involvement, and with Barona estimates in the average range, should generally make 0 or 1 errors on the 20 IF items. Any more than one error, or any errors on the four rarely missed items (57 [c, d, e] and 143 [a; missed by no one in our sample]), suggests the need for further assessment to rule out an underlying cognitive problem, poor effort, or intentional suppression (see Frederick, Crosby, & Wynkoop, 2000, for a discussion of intent and effort on tests of cognition). Importantly, however, before attributing malevolent intent, the clinician must make certain that the subject does not indeed suffer underlying brain pathology. Rogers and Bender (2003) aptly caution "against facile and unwarranted assumptions that suboptimal efforts are always equated with malingering" (p. 109). Of course, the more cognitive items missed in the absence of bonafide CNS illness, the higher the likelihood of an intentionally deviant response style (causal attribution being a clinical decision).

We advocate caution when using the IF scale to characterize effort and intent in brain involved patients. Further, we question the wisdom of administering the entire SIRS to patients with clear and substantial CNS disorders. Our data can most safely be used to characterize or illustrate a person's performance in relation to various patient groups for a trier of fact (e.g., "This person scores at the seventh percentile of patients suffering an encephalopathy despite no basis for this diagnosis").

The combination of cognitive (IF) and psychiatric items on the SIRS has the convenient potential of helping us to better understand the concordance of feigning across these two distinct symptom clusters by malingerers (e.g., a shotgun versus a focused approach to malingering). Future SIRS IF scale research could also focus on cutting scores and classification rates, comparing the IF scale against other more well-established malingering criteria. Persons with specific academic delays and mental retardation should also be included in IF research. Because sensitivity and specificity of the IF scale have not been established, we encourage examiners who use the SIRS to use the IF scale as part of a broader effort and intent assessment strategy.

Disclaimer: The views expressed in this article are not necessarily those of the Federal Bureau of Prisons or of the U.S. Department of Justice.

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author (text rev.).

Barona, A., Reynolds, C. R., & Chastain, R. (1984). A demographically based index of pre-morbid intelligence for the WAIS-R. *Journal of Consulting and Clinical Psychology*, *52*, 885–887.

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

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-mental state. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189–198.

Frederick, R. I., Crosby, R., & Wynkoop, T. F. (2000). Performance curve classification of invalid responding on the validity indicator profile. *Archives of Clinical Neuropsychology*, *15*, 281–300.

Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment*, *6*, 218–224.

Hom, J., & Denney, R. L. (2002). Preface to the special issue on detection of response bias in forensic neuropsychology. Part 1. *Journal of Forensic Neuropsychology*, *2*, xv–xx.

Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, *24*, 1094–1102.

Pankratz, L. (1998). *Patients who deceive*. Springfield, IL: Charles C Thomas.

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

Rogers, R. (1997b). Current status of clinical methods. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (2nd ed., pp. 373–397). New York: Guilford.

Rogers, R., Bagby, R. M., & Dickens, S. E. (1992). *Structured Interview of Reported Symptoms professional manual*. Odessa, FL: Psychological Assessment Resources.

Rogers, R., & Bender, S. D. (2003). Evaluation of malingering and deception. In A. M. Goldstein, & I. B. Weiner (Eds.), *Handbook of Psychology: Forensic Psychology: Vol. 11*,. Hoboken, NJ: Wiley.

Rogers, R., Harrell, E. H., & Liff, C. D. (1993). Feigning neuropsychological impairment: A critical review of methodological and clinical consideration. *Clinical Psychology Review*, *13*, 255–274.

Tan, J. E., Slick, D. J., Strauss, E., & Hultsch, D. F. (2002). How'd they do it? Malingering strategies in symptom validity tests. *The Clinical Neuropsychologist*, *16*, 495–505.

Wynkoop, T. F., Capps, S. C., & Priest, B. J. (1995). Incidence and prevalence of child sexual abuse: A critical review of data collection procedures. *Journal of Child Sexual Abuse*, *4*, 49–66.

Wynkoop, T. F., & Denney, R. L. (1999). Exaggeration of neuropsychological deficit in competency to stand trial. *Journal of Forensic Neuropsychology*, *1*, 29–53.