

# Harry Potter and the Structured Interview of Reported Symptoms?

Author: Gregory DeClue, Ph.D., ABPP (forensic), Sarasota, Florida  
Email: [gregdeclue@me.com](mailto:gregdeclue@me.com)

## Abstract

The developers and publisher of the newly developed Structured Inventory of Reported Symptoms, Second Edition (SIRS-2; Rogers, Sewell, & Gillard, 2010) recommend that evaluators, including forensic evaluators, use the SIRS-2 rather than the original SIRS (Rogers, Bagby, & Dickens, 1992). SIRS-2 classifications of cases as genuine or feigned presentations “are based on empirically validated decision rules rather than norm-referenced interpretations” (Rogers et al., 2010, p. 1). The study purported to support the use of the new classification rules has not been cross-validated, and it has not been subjected to peer review. The authors have failed to provide critical data regarding the classification study. Although “prominent researchers generously made SIRS data available for this [SIRS-2] Professional Manual” (Rogers et al., 2010, p. 73), the test publisher has refused to allow analysis of the data by independent professionals, citing “trade secrets.” As with other proprietary instruments, forensic use of the SIRS-2 appears to be very vulnerable to challenges to its admissibility. Currently, an expert witness who uses the SIRS-2 as a basis for an opinion should inform the trier of fact of important limitations of the instrument.

**Keywords:** SIRS, SIRS-2, Structured Interview of Reported Symptoms, response style, admissibility of psychological testimony, proprietary data, trade secrets

---

As noted in a 2010 review in this journal (Rubenzer, 2010), the very first sentence of the SIRS-2 manual boldly states: “The Structured Interview of Reported Symptoms (SIRS) has been recognized in the last decade as the premier measure for the assessment of feigned mental disorders” (Rogers et al., 2010, p. 1). The SIRS-2 manual continues, “In the clinical forensic arena, the SIRS has been acknowledged by experienced practitioners as a recommended test (i.e., the highest designation for forensic practice; see Lally, 2003) and the most widely used forensic measure of malingering (Archer, Buffington-Vollum, Stredny, & Handel, 2006).”

The second edition of the SIRS, the SIRS-2, was published in February 2010. The developers of the SIRS-2 recommend its use rather than the SIRS, so the SIRS-2 might be expected to replace the SIRS as the premier measure of feigned mental disorders. Prospective users of the SIRS-2 might wish to know, for the same sample, how the SIRS-2 compares to the original SIRS regarding accuracy at distinguishing between feigned versus genuine response patterns, and regarding number of protocols classified (versus deemed indeterminate). That information is being withheld from the public.

Unfortunately, the SIRS-2 authors' and publisher's decision to hold data proprietary makes testimony based on the SIRS-2 vulnerable to being ruled inadmissible in court. That problem could be remedied, easily and promptly.

## Background

### Harry Potter

**Harry Potter** is a series of fantasy novels written by the British author J. K. Rowling about a student at Hogwarts School of Witchcraft and Wizardry. The adolescent wizard came to the attention of forensic psychologists in a decision of the Appeals Court of Massachusetts regarding a sex-offender assessment tool, the Abel Assessment for Sexual Interest (AASI). Ewing (2006) provides a rich discussion of the case and its significance, briefly summarized here. In *Ready v. Commonwealth of Massachusetts*, 824 N.E.2d 474 (2005), the appellate court affirmed a trial court's 2002 exclusion of AASI results in a case in which a convicted sex offender was seeking release from the Massachusetts Treatment Center for Sexually Dangerous Persons.

Gerard Ready had been convicted of multiple counts of raping and sexually assaulting children. He attempted to persuade a state court that he was no longer "sexually dangerous," by offering proof of his performance on the AASI. As described by the trial court, the "AASI is a multi-part assessment device that combines the theory of 'visual reaction time' . . . with a subject's self-reported sexual arousal to slide photograph stimuli. [The AASI] may only be purchased from . . . a for-profit company."

**Science or wizardry?** The State moved to exclude Ready's AASI results from evidence. The trial court applied standards for the admissibility of scientific evidence articulated by the U.S. Supreme Court in *Daubert v. Merrell Dow Pharmaceuticals Inc.*, 509 U.S. 579 (1993). The court held that "the theories and techniques underlying the AASI have not been sufficiently tested for admissibility in this proceeding." After noting that the AASI is a proprietary instrument whose underlying formulas are not released to other professionals, the court wrote, "For all we know, they and their components could be mathematically based, founded upon indisputable empirical research, or simply the magic of young Harry Potter's mixing potions at the Hogwarts School of Witchcraft and Wizardry."

**Repercussions.** The appeals court affirmed exclusion of the AASI, and the Massachusetts Supreme Judicial Court declined to review the appeals court's decision. The courts' opinions in *Ready* exemplify concerns about the admissibility of evidence. A psychologist's opinion testimony based on personal experience could be considered pure-opinion testimony, not subject to an admissibility test such as the *Daubert* test or the standard described in *Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923). However, when the psychologist's testimony is based in whole or in part on the use of psychological tests, the psychologist might be asked to explain such matters as the similarities between the current evaluatee and the study's subjects, links between the test

data and the opinion, and the accuracy of predictions or other interpretations based on test results. When test authors and publishers treat important test data as proprietary trade secrets, the tests are vulnerable to challenges to their admissibility in courts.

## **SIRS**

The original-SIRS manual includes the following (Rogers et al., 1992, p. 1):

The SIRS was designed to evaluate malingering and other forms of dissimulation. The SIRS provides professional users with valuable data on feigning (malingering and factitious disorders with psychological symptoms) and honest responding. In addition, the SIRS offers preliminary evidence of inconsistent responding. In cases of dissimulation, specific SIRS strategies offer insights regarding how the client may be faking. Such insights are likely to be helpful in his or her continued management and treatment.

The original-SIRS manual (pp. 17-18) presents “normative” data for 403 subjects in “very different criterion groups (honest-clinical, honest-nonclinical, suspected malingerers, and simulators) used in making polychotomous decisions about clients (e.g., honest, unreliable, and feigning.” The number of subjects in each of the four criterion groups follows:

- 100 honest-clinical
- 97 honest-nonclinical
- 36 suspected malingerers
- 170 simulators

Tables 16 and 18 (both on page 24) in the original-SIRS manual illustrate decision rules that, for certain patterns of responses, lead to a claimed “likelihood of feigning” of 100.0%, and “likelihood of honest responding” of 95.0%.<sup>1</sup>

## **SIRS-2**

As described in the SIRS-2 manual (p. 69), subsequent research found that some genuine psychiatric patients scored much higher on amplified SIRS-2 strategies than other genuine patient populations. That suggests that a) the original-SIRS classification rules routinely lead to misclassification of some presumed-genuine presentations as feigned and b) the misclassification occurred since the inception of the SIRS and continues to the present, but c) the misclassification was not recognized until cross-validation studies were conducted.

---

<sup>1</sup> Subsequently, the SIRS has been used as a gold standard for research for the MMPI-2, PAI, M-FAST, and SIMS (see Rogers et al., 2010, p. 72 for references and details).

It was recognized that, when generalized beyond the original classification study, the original-SIRS classification rules do not yield classifications as accurate as those claimed in the original manual. Therefore, as the SIRS-2 was developed, four major modifications were implemented in an attempt to more accurately classify presentations (reported symptoms) as honest or feigned (or indeterminate).<sup>2</sup> The authors of the SIRS-2 manual incompletely describe a validation study and report “Utility Estimates for the SIRS-2 Classification of Feigned Mental Disorders” as follows<sup>3</sup>:

- Prevalence of feigning: 31.8%
- False positives: 2.5%
- Sensitivity: .80
- Specificity: .975
- Positive predictive power (PPP): .91
- Negative predictive power (NPP): .91
- Overall correct classification (OCC): .91

In the text accompanying that table, Rogers et al. (2010, p. 89) write, “These findings provide very strong evidence of the SIRS-2 criterion-related validity at the most rigorous level—individual classification.” These are the types of accuracy statistics to which an expert witness would refer when testifying about the basis for an opinion. Unfortunately, information about this key validation study is seriously incomplete, and in some parts inaccurate, in the manual. Indeed, the description in the SIRS-2 manual fails to report some important information that would be expected in any peer-reviewed publication describing such a study.

### **Limitations in the Report of the SIRS-2 Classification Study**

Rogers et al. (2010, pp. 81-82) write, “Criterion-related validity ... directly tests the effectiveness of the SIRS-2 for its primary clinical purpose—specifically, the classification of cases as genuine or feigned mental disorders.” Because the SIRS-2 classification study was not published separately in a peer-reviewed journal, it is important to consider whether the manual includes the quantity and quality of information that would be expected in a manuscript submitted for peer review.

**Description of the Subjects.** When a manuscript regarding an empirical study is submitted for peer review, authors are expected to provide basic information in a way that would allow other researchers to replicate the study and see whether similar results are obtained. Authors are expected to provide a clear, concise description of the sub-

---

<sup>2</sup> Rogers et al. (2010, p. 73) aptly note, “Response styles, such as malingering, are viewed as issue-specific efforts to distort or manipulate clinical presentations. Unlike most measures that attempt to measure stable psychological characteristics or abilities, the SIRS-2 is designed to evaluate, via its detection strategies, two general categories: (a) genuine presentations with circumscribed ranges of scale scores and (b) feigned presentations with marked variations in scale scores.”

<sup>3</sup> These data are presented in Table 4.2 on page 39 of the SIRS-2 manual. The same table is repeated as Table 6.16 on page 89 of the SIRS-2 manual.

jects, including demographics and other characteristics (age, gender, ethnicity, physical or psychiatric diagnoses, etc.) that might influence the outcome of the study.

One key deficiency in the SIRS-2 manual is that the authors fail to adequately identify the subjects. They write (p. 37), “The combined SIRS-2 data set (N = 522) was composed of 314 protocols from the original validation with 208 protocols from the research program at the Timberlawn Mental Health System.” As mentioned above, there were 403 protocols in the original-SIRS classification study. Rogers et al. (2010, p. 37) write, “The original validation included 97 nonclinical controls; they were not used in the SIRS-2 classification.” Subtracting those 97 protocols would yield 306 subjects, not 314. I found no explanation in the SIRS-2 manual for that discrepancy. I requested data in order to independently assess the claims regarding utility statistics for the SIRS-2. The response was that the request should be sent to the publisher because of copyright and proprietary issues.<sup>4</sup> The publisher refused to release the data for independent analysis,<sup>5</sup> despite the fact that “Prominent researchers generously made SIRS data available for this [SIRS-2] Professional Manual” (Rogers et al., 2010, p. 73).

Although the authors of the SIRS-2 manual recommend that the revised instrument should replace “the premier measure for the assessment of feigned mental disorders” (Rogers et al., 2010, p. 1), they do not provide the age, gender, ethnicity, and other characteristics of the subjects. They do not disclose whether there are significant differences in such characteristics between the subjects who provided the presumed-feigned protocols and those who provided the presumed-genuine protocols. They do not provide any basis for treating the patients’ responses as genuine protocols, other than the fact that the subjects were genuinely patients. In addition, the SIRS-2 manual’s authors fail to discuss the fact that most of the presumed-genuine protocols are from test administrations that occurred approximately 20 years after the test administrations for the presumed-feigned protocols.

Inadequacies in the descriptions of the subjects is particularly inexcusable for the SIRS-2 because, “Importantly, [SIRS-2] classifications are based on empirically validated decision rules rather than norm-referenced interpretations” (Rogers et al., 2010, p. 1). That is, classifications (genuine, indeterminate, feigned) are based entirely on differences between groups of subjects in a single study that has not been subjected to peer review.<sup>6</sup> The test publisher clearly presents the SIRS-2 as a forensic instrument: “The

---

<sup>4</sup> Richard Rogers, personal communication, September 2, 2010.

<sup>5</sup> “PAR’s instruments are trade secrets and protected by intellectual property laws including copyright and trade secret laws. I apologize, but in order to protect the integrity, objectivity, security, and copyrights of our psychological data and test materials, we are unable to be supportive of your request.” Vicki King, Executive Assistant to the Chairman and CEO, Psychological Assessment Resources, Inc., personal communication, September 9, 2010.

<sup>6</sup> The classification study compares protocols from five different sources spread over 20 years or so, plus some additional subjects from unpublished data. Page 38 of the original-SIRS manual lists the following sources, and subjects from each of them appear to have been included in the SIRS-2 classification study: “1 = Rogers, Gillis, Dickens, and Bagby (1991), 2 = Rogers, Gillis, and Bagby (1990), 3 = Rogers, Gillis, Bagby, and Monteiro (1991), 4 = [Rogers, Kropp, Bagby, and Dickens (1992)], and 5 = unpublished data

wealth of empirical support established for the measure gives users confidence in the validity of the SIRS-2 for forensic populations. This is particularly important as findings of psychological evaluations are often closely scrutinized during court proceedings.”<sup>7</sup> However, any expert witness who would use the SIRS-2 should be able to testify regarding the extent to which a particular evaluatee is similar to the people included in the SIRS-2 classification study. It is therefore incumbent upon the authors of the SIRS-2 to prepare and submit a research paper for peer review, and for the publishers of the SIRS-2 to allow other professionals to independently analyze the data.

**Representativeness of Subjects.** At the time of the publication of the SIRS-2 manual, there was only one classification study that used the new SIRS-2 classification rules, the one reported in the SIRS-2 manual. For the results of one study to be generalizable to other samples, it is important that the subjects in the original study are representative of the population in question. For the SIRS-2 classification study to be generalizable to the task of distinguishing feigned from genuine presentations of psychopathology, it would be advantageous for the presumed-genuine presentations in the SIRS-2 classification study to come from a representative sample of genuine psychiatric patients. Who are the subjects who provided the presumed-genuine presentations?

Rogers et al. (2010, p. 37) describe the Timberlawn program, from which 208 (approximately two thirds) of the presumed-genuine presentations were obtained, as being “composed primarily of multiply traumatized inpatients who manifested an array of trauma, dissociative, psychotic, and mood symptoms. Their inclusion in the data set broadens the clinical representation for severely impaired genuine patients.” Elsewhere (p. 80) Rogers et al. (2010) mention that “nearly half of the [Timberlawn] participants were diagnosed with dissociative identity disorder.”<sup>8</sup> Thus, approximately two-thirds<sup>9</sup> of the presumed-genuine presentations in the SIRS-2 validation study are from Timberlawn subjects, and over half of those people were diagnosed with dissociative identity disorder (formerly known as multiple personality disorder). Clinical and forensic evaluators are encouraged to consider how similar those subjects are to the populations we assess in our routine practice.

The 207 presumed-feigned presentations come from 36 people who were suspected to be feigning psychopathology at the time they responded to the original SIRS, plus 170 simulators—that is, people who were encouraged to try to feign a mental disorder. According to page 38 of the original-SIRS manual, the 36 suspected feigners included

---

collected for this manual.” Page 69 of the SIRS-2 manual shows that additional subjects were added from Rogers, Payne, Correa, Gillard, & Ross (2009).

<sup>7</sup> See <http://www4.parinc.com/Products/Product.aspx?ProductID=SIRS-2>, accessed November 26, 2010.

<sup>8</sup> The 208 Timberlawn subjects appear to be drawn from the Colin A. Ross Institute for Psychological Trauma, which was founded in 1995 by Colin A. Ross, M.D. The Institute is a private corporation that specializes in the management of psychiatric treatment programs and contracts to provide management and treatment services to Timberlawn Mental Health System, in Dallas, Texas. See more at [http://www.rossinst.com/about\\_ross\\_institute.html](http://www.rossinst.com/about_ross_institute.html)

<sup>9</sup> The exact proportion is unknown, due to the incomplete or inaccurate information about which subjects from the original SIRS classification study were used in the SIRS-2 classification study.

26 forensic evaluatees from Rogers, Gillis, Dickens, and Bagby (1991), plus another 10 evaluatees not included in any prior publications and not described in either of the SIRS manuals. The 170 simulators consisted of 40 community subjects, 70 correctional subjects, and 60 university subjects, drawn from four separate studies, and not described further in the SIRS-2 manual.

The upshot is that, to a significant degree, a clinical evaluator who uses the SIRS-2 classification rules is testing whether the person's responses are more like those of non-patients assessed with instructions to feign psychopathology than those of inpatients diagnosed with dissociative identity disorder and presumed to be offering genuine descriptions of their psychological symptoms.

### **The Problem of Dissociative Identity Disorders**

The Timberlawn study mentioned in the SIRS-2 manual can be viewed as an attempt to cross-validate the original-SIRS classification rules in a larger sample of psychiatric inpatients that included many patients diagnosed with dissociative identity disorder. Prior to the development of the SIRS-2, it was recognized that the original-SIRS classification rules were ineffective at distinguishing between simulators versus inpatients diagnosed with dissociative identity disorder.

In the original-SIRS manual, it is not considered a test flaw that it is necessary to look for external sources of motivation in order to distinguish between various types of feigning, such as malingering versus factitious disorder. Some people have described dissociative identity disorder as a socially constructed phenomenon (contrast Spanos, 1994, 1996, with Gleaves, 1996; see also Lilienfeld & Lynn, 2003), and it may be of some theoretical interest to consider whether typical patients diagnosed with dissociative disorder should be considered false positives if classification rules would label their symptom descriptions as somehow fake. In forensic practice, though, it would be useful to be able to reliably distinguish between people who are honestly reporting what they believe their symptoms to be (whatever the etiology might be), versus people who are deliberately exaggerating or lying about their symptoms.

A 2006 study using the original SIRS reported, "Test responses from 20 dissociative identity disorder (DID) patients are compared to those of 43 well informed and motivated DID simulators. Both the simulators and DID patients endorsed such a high number of symptoms that their average overall scores would typically be interpreted as indicative of feigning. The simulators' mean scores were significantly higher than those of the DID patients on only four out of 13 scales" (Brand, McNary, Loewenstein, Kolos, & Barr, 2006, p. 63).

In the 2009 study that directly contributed to the development of the SIRS-2, the researchers "recruited traumatized patients"<sup>10</sup> for a within-subjects simulation design in

---

<sup>10</sup> It appears that the terms "traumatized patients" and "patients diagnosed with dissociative identity disorders" are used at least somewhat interchangeably. Some readers may be interested in Lilienfeld

which we asked feigners to convincingly portray themselves as examinees claiming total disabilities. When compared to standard instructions, feigned presentations produced substantial effect sizes. Although the standard SIRS classifications produced moderately high sensitivities . . . , the false-positive rates were problematic. To minimize false-positives, we constructed a Trauma Index (TI) from 3 primary SIRS scales that appeared unaffected by severe trauma. Implementation of the TI substantially reduced false-positive rates” (Rogers, Payne, Correa, Gillard, & Ross, 2009, p. 429).

In considering how to avoid classifying protocols from DID patients as feigned, the developers of the SIRS-2 considered advising against using the instrument for certain subjects,<sup>11</sup> but decided instead to develop new classification rules that could be applied for any person tested.<sup>12</sup>

### **Enhanced Accuracy of SIRS-2 Classification? At What Cost?**

Development and use of classification rules inevitably involve trade-offs. For example, developers of a classification test may be able to decrease false positives in a way that also decreases true positives and increases the number of people considered indeterminate. In order to make such trade-offs explicit, test developers are encouraged to publicly present not only the classification rules/recommendations, but also the data that led the developers to those recommendations. For example, Hart, Webster and Menzies (1993, p. 698) recommend, “To help prevent future misinterpretations and to facilitate inspection and (re-)analysis by readers, we recommend that journal editors require authors to report in their manuscript the raw data for any 2 X 2 analyses. Such data are easily presented in the form of text or tables.”<sup>13</sup>

It might be expected that the manual for the SIRS-2 would show the impact of using the SIRS-2 classification rules versus the original-SIRS classification rules on the number of accurate classifications, inaccurate classifications, and cases deemed indeterminate. Because the SIRS-2 manual does not report that, I requested the data from the test author, and then from the publisher. As mentioned above, those requests were not fulfilled, reportedly to protect trade secrets. The following is an attempt to reconstruct the parts of that analysis that are publicly available, and to show which parts are currently being held from public view.<sup>14</sup>

---

and Lynn’s (2003) discussion of the extent to which such patients’ accounts of childhood trauma have been independently verified.

<sup>11</sup> “One consideration was to potentially limit the use of the SIRS-2 for trauma patients, especially those who report childhood trauma and dissociative symptoms” (Rogers et al., 2010, p. 69).

<sup>12</sup> “An early decision was effectuated to have one set of rules that could be applied broadly rather than attempting different decision rules based on clinical presentation (e.g., dissociative episode)” (Rogers et al., 2010, p. 69). The authors of the SIRS-2 manual do not discuss how they ruled out the possibility that a significant proportion of the dissociative patients’ responses were feigned or otherwise unreliable.

<sup>13</sup> Hart et al.’s recommendation was offered in the context of risk assessment, but it is equally applicable here.

<sup>14</sup> I provided a pre-publication copy of this article to Psychological Assessment Resources (PAR; publisher of the SIRS-2) on December 17, 2010, to allow a check for factual errors. PAR did not respond.



**1. Traditional 2 x 2 Classification Tables.**<sup>15</sup> When using a test to classify a person's responses as feigned or genuine, four outcomes can occur:

- True Positive: We classify a protocol as feigned, and it was feigned.
- True Negative: We classify a protocol as genuine, and it was genuine.
- False Positive: We classify a protocol as feigned, but it was genuine.
- False Negative: We classify a protocol as genuine, but it was feigned.

These values are typically presented in a 2 X 2 table such as Table 1.<sup>16</sup>

Table 1  
A Model 2 X 2 Table for a Binary Classification Test

	Presumed Feigned	Presumed Genuine	
Classified as Feigned	<b>True Positive (TP)</b>	<b>False Positive (FP)</b>	→ Positive Predictive Value (PPV)
Classified as Genuine	<b>False Negative (FN)</b>	<b>True Negative (TN)</b>	→ Negative Predictive Value (NPV)
	↓ True Positive Rate (TPR)	↓ False Positive Rate (FPR)	

For a particular test, research data are plugged into the center four cells (TP, FP, FN, and TN). Then standard formulas can be used to calculate the following, all of which help users understand how accurate the test is at predicting recidivism.

- sample base rate (BR)
- true positive rate (TPR), also called sensitivity
- false positive rate (FPR)
- specificity
- positive predictive value (PPV)
- negative predictive value (NPV)

The sample *base rate (BR)* is the proportion of people with a condition. Here, it is the number of protocols in the sample that are presumed to have been feigned.

<sup>15</sup> A more-detailed discussion is provided in this journal in Campbell & DeClue (2010).

<sup>16</sup> See, for example, Clinical Calculator 1 at <http://faculty.vassar.edu/lowry/VassarStats.html>

The *true positive rate (TPR)*, also called *sensitivity*, is the proportion of actual positives that are identified. Here, it is the proportion of protocols that were correctly classified as feigned. It is calculated as  $TP / (TP + FN)$ .

The *false positive rate (FPR)* is the proportion of protocols incorrectly classified as genuine. It is calculated as  $FP / (FP + TN)$ .

*Specificity* is the proportion of genuine protocols correctly identified as genuine. It is calculated as  $TN / (TN + FP)$ .

The *positive predictive value (PPV)* is the probability that a protocol was feigned, given that it met rules for being classified as feigned. It is calculated as  $TP / (TP + FP)$ .

The *negative predictive value (NPV)* is the probability that a protocol was genuine, given that it met rules for being classified as genuine. It is calculated as  $TN / (FN + TN)$ .

**2. A 2 x 2 Classification Table for the SIRS-2.** It is possible to construct a traditional 2 x 2 classification table from the information provided on page 38 of the SIRS-2 manual. Table 2 presents data for 402 of the 522 subjects in the SIRS-2 classification study.

Table 2  
SIRS-2 Classification Sample Scored Using SIRS-2 Rules: Classified Cases Only

	Presumed Feigned	Presumed Genuine	
Classified as Feigned	102 (TP)	10 (FP)	112
Classified as Genuine	26 (FN)	264 (TN)	290
	128	274	402

Sample BR = .318, TPR (Sensitivity) = .800, FPR = .036, Specificity = .964  
Overall Accuracy at this BR = .910, PPV at this BR = .911, NPV at this BR = .910

In the SIRS-2 manual, incorrect numbers are provided for FPR and specificity in their Table 4.2 on page 39.<sup>17</sup> (The same table is repeated as their Table 6.16 on page 89 of the SIRS-2 manual.) In addition to being inaccurate in parts, the reported test utilities for the SIRS-2 classification are incomplete in at least two ways. First, they ignore the fact that 120 of the 522 protocols (23%) were classified as indeterminate. That is, more protocols were classified as indeterminate than as feigned, and those classified as indeterminate were ignored when considering how accurate the SIRS-2 classification rules are. Second, the SIRS-2 manual does not provide data regarding how the subjects in the SIRS-2 classification study would be classified if the original-SIRS rules were used.

<sup>17</sup> In the SIRS-2 manual, FPR is reported as 2.5%, and specificity is reported as .975.

**3. A 3 x 2 Classification Table for the SIRS-2.** Rogers et al. (2010, pp. 37, 38) report that there were 522 protocols in the classification study, and that 402 were classified as either feigned or genuine. We can begin to construct a more complete 3 x 2 table, as shown in Table 3a.

Table 3a  
SIRS-2 Classification Sample Scored Using SIRS-2 Rules

	Presumed Feigned	Presumed Genuine	
Classified as Feigned	102	10	112
Classified as Indeterminate	?	?	120
Classified as Genuine	26	264	290
	?	?	522

If we assume that all 207 presumed-feigned protocols from the original-SIRS classification study were included in the SIRS-2 classification study, the results would be as shown in Table 3b.

Table 3b  
SIRS-2 Classification Sample Scored Using SIRS-2 Rules

	Presumed Feigned	Presumed Genuine	
Classified as Feigned	102	10	112
Classified as Indeterminate	79	41	120
Classified as Genuine	26	264	290
	207	315	522

Some important differences arise when we consider Table 3b versus Table 2. Test utilities derived by going across the rows in the tables remain the same. PPV is still .911, and NPV is still .910. That is, 91% of protocols classified as feigned are from the “presumed feigned” group, and 91% of protocols classified as genuine are from the “presumed genuine” group. However, test utilities derived by going down the columns in the tables are not the same.

For Table 2 (the 2 x 2 contingency table), TPR, also called sensitivity, is reported as .800 and is described as the proportion of presumed-feigned protocols that are correctly classified. However, when we consider all of the subjects (including those classified as indeterminate), as shown in Table 3b, we realize that 102 of 207 presumed-feigned protocols (49%) are correctly classified using the SIRS-2 rules. That is, although the

SIRS-2 manual claims a specificity of .80, the odds are actually only about 50-50 that a feigned protocol in that sample was correctly identified as such.

**4. A 3 x 2 Classification Table for the SIRS-2 Classification Sample, Using the Original-SIRS Classification Rules.** To what extent do the SIRS-2 classification rules improve the accuracy of classification? That study has been done, but the data have not been reported. As mentioned previously, the publisher is withholding the data, reportedly to protect trade secrets. Readers (including potential test users) should be able to directly compare the accuracies of the original SIRS versus the SIRS-2. That would be possible if the authors and/or publisher of the SIRS-2 would release data sufficient to fill in Table 4. Then readers (including potential test users) could see the trade-offs in accuracy, number of cases deemed indeterminate, etc., that come from using SIRS-2 rather than original-SIRS classification rules for these subjects.

Table 4  
SIRS-2 Sample Scored Using Original-SIRS Rules

	Presumed Feigned	Presumed Genuine	
Classified as Feigned	?	?	?
Classified as Indeterminate	?	?	?
Classified as Genuine	?	?	?
	207	315	522

**5. A 3 x 2 Classification Table for the Original-SIRS Classification Sample, Using the SIRS-2 Classification Rules.** Data to complete Table 4 (above) would allow readers (and potential test users) to compare original-SIRS versus SIRS-2 accuracy rates for the SIRS-2 classification sample. Recall that the presumed-genuine protocols in the SIRS-2 classification sample were provided by an unusual group of patients, with more than one third of them diagnosed with dissociative identity disorder. The developers of the SIRS-2 assumed that all of those protocols were genuine reports of symptoms, and they developed the SIRS-2 classification rules in order to decrease the number of those protocols that would be classified as feigned. What effect does that have in other patient populations? Are more feigned protocols classified as indeterminate? If so, how many more?

Data to address those questions are readily available to the SIRS-2 developers, but have not been reported. Data from the 100 presumed-genuine<sup>18</sup> and 207 presumed-feigned subjects in the original-SIRS classification sample could be used to fill in Tables 5 and 6.

<sup>18</sup> The 100 presumed-genuine protocols from clinical subjects in the four samples included in the original-SIRS classification sample would be included (see page 38 of the original-SIRS manual).

Table 5  
Original-SIRS Sample Scored Using SIRS-2 Rules

	Presumed Feigned	Presumed Genuine	
Classified as Feigned	?	?	?
Classified as Indeterminate	?	?	?
Classified as Genuine	?	?	?
	207	100	307

Table 6  
Original-SIRS Sample Scored Using Original-SIRS Rules

	Presumed Feigned	Presumed Genuine	
Classified as Feigned	?	?	?
Classified as Indeterminate	?	?	?
Classified as Genuine	?	?	?
	207	100	307

### Proprietary Data

Reliance on proprietary data is anathema to modern science and, in particular, to science that would inform legal opinions. For example, the Editorial and Publishing Policies of *Public Library of Science One* include, “The conclusions of a study must not be dependent solely on the analysis of proprietary data. If proprietary data were used to reach a conclusion, and the authors are unwilling or unable to make these data public, then the paper must include an analysis of public data that validates the conclusions so that others can reproduce the analysis and build on the findings.”<sup>19</sup>

Although the classification rules for the SIRS-2 are clearly and publicly described in the manual, the claimed test utilities (e.g., sensitivity, specificity, false positives) are based on data that are incompletely described, include inaccuracies, and have not been made available for independent analysis by other professionals. This fails to comply with Hart et al.’s (1993) recommendation, and is reminiscent of the Massachusetts appellate court’s decision to exclude testimony based on a proprietary instrument (*Ready v. Commonwealth*, 2005).

<sup>19</sup> <http://www.plosone.org/static/policies.action>

## Discussion

It is significant that the original-SIRS classification rules fail to discriminate between people diagnosed with dissociative identity disorder versus well-informed and motivated simulators. The developers of the SIRS-2 recognized that something needed to be done lest certain people who honestly report what they believe their psychological symptoms to be would be misclassified as people who were deliberately exaggerating or feigning symptoms. Development of new classification rules for everyone is one reasonable, potential approach to the problem, and the newly developed SIRS-2 classification rules *might* be the best solution. However, because the SIRS-2 authors and publisher are keeping essential data secret, it is impossible for independent professionals to evaluate the test authors' and publisher's claims.

The SIRS-2 classification rules are not based on peer-reviewed research. When the test developers noticed that the original-SIRS classification rules did not cross-validate well with a specialized population, they revised the classification rules for everyone. The new classification rules have not been subjected to an independent cross-validation (separate from the subjects used in developing the SIRS-2 classification rules), so at this point it is unknown whether the SIRS-2 classification rules will lead to more accurate classifications than the original-SIRS classification rules if and when the two sets of rules are compared with new subjects.

The developers of the SIRS-2 should prepare and submit a manuscript for peer review, providing all of the information that would be expected for consideration of any study that claims to provide evidence sufficient to warrant abandoning "the premier measure for the assessment of feigned mental disorders" and replacing it with a new test. That is especially true when the new test is published by a for-profit company. Professionals should be just as skeptical as they would be if the following were not true:

1. The first author of both instruments is the same person.
2. The same company publishes both instruments.
3. The two instruments overlap considerably in content and administration.

For a forensic professional, the insufficiency of these facts becomes readily apparent when one imagines oneself on a witness stand testifying that he or she chose to use a particular instrument for reasons such as the following:

1. The test developer recommended that I buy and use the new test.
2. The test publisher recommended that I buy and use the new test.
3. The new test looks a lot like the old test, and it's new!

It is not clear why the test developers and publisher are treating the data as trade secrets. Because both the SIRS and the SIRS-2 are hand scored, the ongoing per-use costs for the tests come from purchases of additional test booklets. Release of the currently withheld "trade-secret" data would not affect ongoing costs (to the user) or income

(to the publisher). The only financial impact of releasing the secret data would accrue if potential test users who understood the instrument better would decide not to buy and use the instrument. Such considerations fuel skepticism regarding proprietary instruments, not just by courts but also by discerning clinicians.

Whatever the motivation of the publisher to treat essential data regarding the SIRS-2 validation study as “trade secrets,” the impact for forensic professionals is that the SIRS-2 appears to be very vulnerable to challenges regarding its admissibility in court. An attorney could argue, “Your Honor, in *Ready* the Court wrote, ‘For all we know, they and their components could be mathematically based, founded upon indisputable empirical research, or simply the magic of young Harry Potter’s mixing potions at the Hogwarts School of Witchcraft and Wizardry.’ That was true for *Harry Potter and the Abel Assessment for Sexual Interest*. It is also true for *Harry Potter and the Structured Interview of Reported Symptoms*. And it will be true for any case in which an expert wants a Court to base a decision on secret science.”

No wizardry is required to avoid this fate. All that is needed is for the developers of the SIRS-2 to develop and submit a manuscript regarding the SIRS-2 classification study that is sufficiently comprehensive to survive peer review, and for the publisher to allow independent professionals to analyze the data without impediment.

### **Current Testimony**

In the meantime, an expert witness who uses the SIRS-2 as a basis for an opinion should inform the trier of fact of important limitations of the instrument. These include:

1. The SIRS-2 classification rules (e.g., feigned, genuine) are dependant solely on the analysis of proprietary data, not available for verification by the scientific and professional communities.
2. The SIRS-2 classification rules have not been cross-validated. No publicly available data validate the authors’ claims.
3. The values for FPR and specificity reported in the SIRS-2 manual appear to be incorrect.
4. The reported sensitivity for the SIRS-2 is misleading in that it is calculated as if there were no indeterminate cases. Within the SIRS-2 classification study, approximately half of the presumed-feigned protocols were identified as feigned, not 80% as reported in the manual.
5. Subjects in the classification study are drawn from an amalgam of studies spanning approximately 20 years, and subject characteristics (e.g., age, gender, diagnosis) are unreported for some of the subjects. It may be difficult or impossible for an evaluator to testify regarding how similar the subject of a forensic case is to subjects in some of the samples within the SIRS-2 classification study.

## References

- Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V., & Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of Personality Assessment, 87*, 84–94.
- Brand, B. L., McNary, S. W., Loewenstein, R. J., Kolos, A. C., & Barr, S. R. (2006). Assessment of genuine and simulated dissociative identity disorder on the Structured Interview of Reported Symptoms. *Journal of Trauma and Dissociation, 7*, 63-85.
- Campbell, T. W., & DeClue, G. (2010). Maximizing predictive accuracy in sexually violent predator evaluations. *Open Access Journal of Forensic Psychology, 2*, 148-232.
- Daubert v. Merrell Dow Pharmaceuticals Inc.*, 509 U.S. 579 (1993).
- Ewing, C. P. (2006). Testing tool in question. *Monitor on Psychology, 37*(1), 61. Retrieved November 26, 2010, from <http://www.apa.org/monitor/jan06/jn.aspx>.
- Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923).
- Gleaves, D. H. (1996). The sociocognitive model of dissociative identity disorder: A reexamination of the evidence. *Psychological Bulletin, 120*(1), 42-59.
- Hart, S. J., Webster, C. D., & Menzies, R. J. (1993). A note on portraying the accuracy of violence predictions. *Law and Human Behavior, 17*, 695-700.
- Lally, S. J. (2003). What tests are acceptable for use in forensic evaluations? A survey of experts. *Professional Psychology: Research and Practice, 5*, 491–498.
- Lilienfeld, S. O., & Lynn, J. M. (2003). Dissociative identity disorder: Multiple personalities, multiple controversies. In S. O. Lilienfeld, S. J. Lynn, & J. M. Lohr, *Science and pseudoscience in clinical psychology*. New York: Guilford.
- Ready v. Commonwealth of Massachusetts*, 824 N.E.2d 474 (2005).
- Rogers, R., Bagby, R. M., & Dickens, S. E. (1992). *Structured Interview of Reported Symptoms (SIRS), Professional manual*. Lutz, FL: Psychological Assessment Resources, Inc.
- Rogers, R., Kropp, R., Bagby, R. M., & Dickens, S. E. (1992). Faking specific disorders: A study of the Structure Interview of Reported Symptoms (SIRS). *Journal of Clinical Psychology, 48*, 643-647.



- Rogers, R., Gillis, J. R., & Bagby, R. M. (1990). Cross-validation of the SIRS 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 SIRS: A study of coached and uncoached simulators. *Psychological Assessment: A Journal of Clinical and Consulting Psychology*, 3, 673-677.
- Rogers, R., Gillis, J. R., Dickens, S. E., & Bagby, R. M. (1991). Standardized assessment of malingering: Validation of the SIRS. *Psychological Assessment: A Journal of Clinical and Consulting Psychology*, 3, 89-96.
- Rogers, R., Payne, J. W., Correa, A. A., Gillard, N. D., Ross, C. A. (2009). A study of the SIRS with severely traumatized patients. *Journal of Personality Assessment*, 91, 429-438.
- Rogers, R., Sewell, K. W., & Gillard, N. D. (2010). *Structured Interview of Reported Symptoms (SIRS), 2<sup>nd</sup> Edition, professional manual*. Lutz, FL: Psychological Assessment Resources, Inc.
- Rubenzler, S. (2010). Review of the Structured Interview of Reported Symptoms-2 (SIRS-2). *Open Access Journal of Forensic Psychology*, 2, 273-286.
- Spanos, N. P. (1994). Multiple identity enactments and multiple personality disorder: A sociocognitive perspective. *Psychological Bulletin*, 116, 143-165.
- Spanos, N. P. (1996). *Multiple identities and false memories*. Washington, D.C.: American Psychological Association.

Received December 13, 2010; revision submitted January 18, 2011; accepted January 19, 2011