WestVirginiaUniversity
THE RESEARCH REPOSITORY @ WVU

Graduate Theses, Dissertations, and Problem Reports

2007

Utility of the Structured Inventory of Malingered Symptomatology (SIMS) and the Assessment of Depression Inventory (ADI) in screening for malingering among disability seeking outpatients

Carl B. Clegg
West Virginia University

Follow this and additional works at: https://researchrepository.wvu.edu/etd

Utility of the Structured Inventory of Malingered Symptomatology (SIMS) and the Assessment of Depression

Inventory (ADI) in Screening for Malingering Among Disability Seeking Outpatients

Carl B. Clegg

Thesis submitted to the
College of Arts and Sciences
at West Virginia University
in partial fulfillment of the requirements
for the degree of

Master of Science
in
Clinical Psychology

William Fremouw, Ph. D., Chair
Amy Fiske, Ph. D.
Kent Parker, Ph.D.

Department of Psychology

Morgantown, West Virginia
2007

Keywords: malingering, disability, assessment, depression

Abstract

Utility of the Structured Inventory of Malingered Symptomatology (SIMS) and the Assessment of Depression

Inventory (ADI) in Screening for Malingering Among Disability Seeking Outpatients

Carl B. Clegg

Sixty-four individuals undergoing a social security disability evaluation were administered the Structured Interview of Reported Symptoms (SIRS), Structured Inventory of Malingered Symptomatology (SIMS), and Assessment of Depression Inventory (ADI). Individual results were classified as honest or suspected malingering based on SIRS scores. In addition, 63 individuals from the community were randomly assigned to complete the SIMS and the ADI honestly or as if they were malingering depression. Both malingering groups had significantly higher mean scores on the SIMS Total and ADI Feigning scales than both honest groups. The scores of the malingering groups did not significantly differ. In the clinical sample, previous recommended SIMS Total cut-scores (> 14 or > 16) had excellent sensitivity, but low specificity. Conversely, the recommended ADI Feigning cut-score (> 13) had excellent specificity, but low sensitivity. Higher and lower cut-scores on the SIMS Total and the ADI Feigning scale, respectively, may improve their utility in screening for malingering among disability seeking outpatients.

Acknowledgements

The successful completion of this study would not have been possible without the support and guidance of Dr. William Fremouw, my committee chair. I would also like to thank my committee members, Dr. Amy Fiske and Dr. Kent Parker, for their time and effort in reviewing my manuscript and for their suggestions. In addition, my undergraduate research assistants, Alyssa Giron and Johanna Wendel, and fellow graduate student, Tracy Thomas, also deserve special thanks for assisting in data collection.

Table of Contents

List of Tables

List of Figures

Utility of the Structured Inventory of Malingered Symptomatology (SIMS) and the Assessment of Depression

Inventory (ADI) in Screening for Malingering Among Disability Seeking Outpatients

The need to assess for potential malingering among disability claimants is clear, but exactly how such an assessment should be conducted is not. The *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV-TR; American Psychiatric Association [APA], 2000) defines malingering as "the intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by external incentives" (p. 739). This definition differentiates the term malingering from feigning by identifying a motive. It also clarifies the distinction between malingering and conversion or other somatoform disorders, where the exaggerated symptoms are not intentionally produced, and from factitious disorder, where the motivation is internal (i.e., to assume the sick role). It states that malingering should be strongly suspected if two or more of the following are noted: medicolegal context of presentation, marked discrepancy between the person's claimed stress and disability and the objective findings, lack of cooperation during the diagnostic evaluation and in complying with the prescribed treatment regimen, and the presence of Antisocial Personality Disorder. These criteria are insufficient to make a formal classification of malingering, however, and have been shown to be inadequate even as screening guidelines (Widows & Smith, 2005).

The estimates for malingering range widely from 1% to 50% depending on context (Rogers & Salekin, 1998). Malingering in forensic settings is estimated to be around 15-17 % (Rogers & Bender, 2003) and estimates specifically for disability claims are even higher. For instance, a survey of 131 members of the American Board of Clinical Neuropsychology suggests malingering in 30% of disability claims (Mittenberg, Patton, Canyock, & Condit, 2002). Also, Miller (2001) reported that 6 of 16, or 37.5%, of disability seekers were malingering based on Structured Interview of Reported Symptoms (SIRS; Rogers, Hicks, & Bagby, 1992) scores.

Without reliable and generally accepted diagnostic criteria, clinicians vary in how they assess malingering. Some of this variability might be due to different theoretical views about the etiology of malingering. For example, Rogers (1992) described three explanatory models of malingering: pathogenic, criminological, and adaptational. The pathogenic model is primarily psychodynamic and considers the fabrication or exaggeration of symptoms as an unconscious attempt to ward off impending psychoses. The criminological model basically asserts that someone must be overtly uncooperative and have Antisocial Personality Disorder in order to malinger. The idea that malingerers must have an impending or existing mental disorder has not been demonstrated empirically, however, so

both of these models have generally fallen into disfavor. Currently, most clinicians subscribe to the adaptational model which considers malingering to be a conscious choice made by individuals who perceive it as the best way to further their self-interests in an adversarial situation. Although a substantial percentage of malingerers may also have a genuine mental disorder, the adaptational model can also explain malingering in individuals who do not.

Variability in assessment strategies is probably mostly due to differing beliefs regarding detection. There are three main approaches that forensic psychologists take when evaluating response styles to determine the likelihood of malingering (Rogers & Bender, 2003). The first is the *intuitional* approach which assumes that clinical acumen alone is sufficient to detect malingering. Although it lacks empirical validation, many clinicians still rely on this approach. This may be due in part to the mistaken belief that malingering is very rare and/or easily detectable. However, as reported earlier, malingering is not uncommon. Also, Rogers (2001) noted that malingering may have gone undetected in disability cases simply because the feigning was not done in an obvious manner.

The second, and most commonly used, is the *standard* approach which is the assessment of malingering based on traditional tests that have built-in validity scales (e.g, MMPI-2, PAI) that address different types of dissimulation. Although some of these embedded scales appear to be useful in certain contexts, there is no consensus about their cut-scores (the value chosen if scores are above or below to be indicative of dissimulation and scores in the opposite direction to be indicative of honesty) and generalizability. For instance, in a meta-analytic study of the MMPI- 2, Rogers, Sewell, and Salekin (1994) found reported cut-scores on the *F* scale for the classification of feigning ranging from > 8 to > 30. Also, very little research on these feigning scales has been conducted with suspected malingerers or minority populations (Rogers, 2001).

The third is the *specialized* approach which supplements traditional testing with measures specifically designed to detect feigning in response style. This approach has the greatest classificatory accuracy and, considering the potentially devastating consequences of mislabeling a genuine patient as a feigner, Rogers & Bender (2001) recommend it as the necessary model in the determination of malingering.

Although most clinicians who evaluate disability claims are not naïve to potential malingering, they may not utilize the specialized approach because of practical constraints. For instance, currently the most extensively validated measure designed to detect feigning in response style is the Structured Inventory of Reported Symptoms (SIRS; Rogers et al., 1992), but a limitation of the SIRS in clinical practice is that it is a structured interview that requires nearly an hour to administer. This is beyond the time available for most standard diagnostic assessments.

Evaluating every disability claimant in an in-depth, malingering-specific manner may not be a feasible expenditure of a clinician's limited time. Essentially, there is a trade-off between greater classificatory accuracy and practicality. To resolve this dilemma, a number of malingering specific self-report screening measures have been developed in recent years that can be completed by the client in 5 -15 minutes without the clinician being present. The utility of these measures are still in question, but two that show particular promise with a disability seeking population are the Structured Inventory of Malingered Symptomatology (SIMS; Widows & Smith, 2005) and the Assessment of Depression Inventory (ADI; Mogge & LePage, 2004).

The Structured Inventory of Malingered Symptomatology (SIMS; Widows & Smith, 2005) was designed to be a self-report screening instrument for malingering. It is a 75-item true-false test that can be administered in about 15 minutes. The SIMS has five subscales: Affective Disorders (AF), Amnestic Disorders (AM), Low Intelligence (LI), Neurologic Impairment (NI), and Psychosis (P). Each subscale contains 15 items designed to identify malingering of a particular type of disability. A total score is also calculated as the sum of raw scores from all five scales. A total score greater than 14 is indicative of possible malingering.

Nine studies have been done to evaluate the validity of the SIMS. Six of these were simulation studies, two used known-group comparisons, and one had both a simulation and a known-group component. In their initial validation study, Smith and Burger (1993) randomly assigned 238 undergraduates to either a control group (honest responders) or one of six experimental conditions where they were instructed to simulate one of the following: psychosis (P), amnesia (A), depression (D), low intelligence (LI), neurologic impairment (NI), or faking bad (FB). All of the malingering groups had significantly higher scores on every subscale than the honest group. In addition, every group had the highest mean score on their respective subscale. Although those in the P group had similarly high scores on the NI and A scales, and the NI group performed similarly on the P scale. These results suggested fairly good construct validity for the SIMS in differentiating between honest and feigned responding, overall, as well as differentiating between specific types of feigned symptoms.

Rogers, Hinds, and Sewell (1996) examined the utility of the SIMS with an adolescent population. Participants, ages 14 to 17 and in court referred residential treatment, filled out the SIMS under both honest and simulated malingering instructional sets. For the malingering condition, participants were randomly assigned to present with one of three conditions: schizophrenia, major depression, or generalized anxiety disorder. They found that a cut-score of > 16 yielded a positive predictive power (PPP; the percentage of participants identified by the

SIMS as malingering who actually were instructed to malinger) of .87, and suggested this higher cut-score may be more useful than > 14.

Edens, Otto, and Dwyer (1999) used a similar design as Rogers et al. (1996) with undergraduates, except the three simulated conditions were: psychosis, depression, or cognitive impairment. Their results showed that a SIMS total score (using a cut-score of > 14) is more useful than the subscales in differentiating honest from feigned responding. They also found that genuine psychopathology (based on Symptom Checklist-90-Revised [SCL-90-R; Derogatis, 1975] scores) tends to increase SIMS scores, but typically not enough to suggest feigning.

Merckelbach and Smith (2003) examined the psychometric properties of a Dutch translation of the SIMS. In one sample, they randomly assigned 82 undergraduates to either an honest group or one of three simulation groups: amnesia, schizophrenia, or neurological problems. They also randomly recruited 10 psychiatric inpatient volunteers to respond honestly. They found significant differences between honest and feigned responding groups on the SIMS total score and the relevant subscales. In addition, they found that a cut-score of > 16 was the most effective with a PPP of .90 and a negative predictive power (NPP; the percentage of participants identified by the SIMS as honest who actually were instructed to answer honestly) of .98.

Poythress, Edens, and Watkins (2001) were the first to study the SIMS with suspected malingerers. They examined four groups: two from a prison's general population (honest vs. simulated) and two from the prison's mental health unit (honest vs. residents diagnosed as suspected malingerers based on SIRS scores). There were 26 suspected malingerers and 30 participants in the other three groups. A SIMS total score of > 14 had the best overall discrimination, misclassifying only 10.3% of simulated malingerers and 15.4% of clinical malingerers.

Heinze and Pursich (2001) used suspected malingerers as well, but without a comparison group. Fifty-seven men who were previously found to be malingering, based on a clinical evaluation for competency to stand trial, were administered the SIMS along with several other measures used to detect malingering (e.g., SIRS). Their scores were compared to a hypothetical clinical sample and a hypothetical malingering sample, derived from examining the manuals and the literature on clinical samples (using the most conservative means). Participants scored significantly higher than the hypothetical clinical group on all SIMS scales, but scored lower than the hypothetical malingering group on all scales, and significantly lower on the total score and the P subscale. The lack of true comparison groups, however, is a severe limitation of this study.

The most recent known-groups study on the SIMS was conducted by Lewis, Simcox, and Berry (2002). They administered the SIRS, MMPI-2, and the SIMS to 64 men undergoing pretrial forensic assessments at two Federal Bureau of Prisons facilities. They found that using a cut-score of > 16 on the SIMS had a NPP of 1.00 for classifying non-feigners identified by the SIRS. However, this same cut-score had a PPP of only .54. Thus, reinforcing the use of the SIMS only as a screening measure.

Rogers, Jackson, and Kaminski (2005) used the SIMS in a simulation study that sought to differentiate factitious feigning from malingering. Sixty-five doctoral students in psychology were assigned to one of four conditions: factitious-dependent, factitious-demanding, malingering, or honest (control). Mean total scores on the SIMS were significantly higher for simulated malingerers than for controls. However, on the subscales, only AF (affective) scores were significantly different between these groups.

Lastly, Jelicic, Hessels, and Merckelbach (2006) conducted a simulation study that randomly assigned participants to one of four conditions where they were either instructed to answer honestly (controls), malinger psychosis with no further information given (naïve malingerers), malinger with some information about psychosis given (informed malingerers), or malinger with information about psychosis and a warning not to exaggerate symptoms (coached malingerers). Their findings suggest that coached malingerers are no more successful than uncoached malingerers in avoiding detection on the SIMS. However, their "coaching" was only providing several minutes of general information about the core symptoms of psychosis. Therefore, the susceptibility of the SIMS to extensive or more specific coaching remains untested.

Recently, Geraerts, Jelicic, and Merckelbach (2006) used the SIMS in a study examining the possibility of symptom overreporting among individuals with recovered memories of childhood sexual abuse. The validity of the SIMS was not examined in this study, rather, it was used as a dependant measure to indicate symptom overreporting. They found that people with recovered memories of childhood sexual abuse are no more likely to overreport symptoms than people who have had continuous memories of childhood sexual abuse (i.e., they did not obtain significantly higher SIMS scores). However, 16% of the participants with continuous memories of childhood sexual abuse scored above the feigning cut-score (> 16) while only .01% participants in the control group (individuals with no childhood sexual abuse) did.

These results were puzzling to the researchers because the participants were informed the research data could not be used for legal or medical purposes. Why, then, would the participants with continuous memories of

childhood sexual abuse feign symptoms? They concluded the most likely explanation for these results was the wish of the participants to make a believable impression on them (the researchers). However, congruent with the findings by Edens et al. (1999), these results could also be due to the SIMS sensitivity to genuine psychopathology and not just feigning.

In the SIMS professional manual, Widows and Smith (2005) summarize that the total score more reliably differentiates between malingering and honest responding than the subscales. Although the subscales may have utility in identifying specific types of disorders (e.g., memory impairment, psychosis). They report alphas coefficients ranging from .82 (P scale) to .88 (Total score) for internal consistency. The total score also has adequate test-retest reliability ($r = .72$). Even though some researchers found a total cut score of > 16 having the greatest utility (Rogers et al., 1996; Lewis et al., 2002) they recommend a cut score of > 14 because the SIMS is intended to be a screening measure and not a diagnostic instrument. Therefore NPP is of greater importance than PPP because false positives will likely be rectified in the more thorough follow-up testing (e.g., SIRS) that would be warranted.

The Assessment of Depression Inventory is a 39-item, self-report measure that takes about five minutes to complete. It assesses the presence of real depressive symptomology and contains validity scales for feigning, random, and unreliable responding. Respondents are asked how frequently in the last two weeks they have experienced the feelings and reactions defined in the 39 items. They rate each item on a four-point scale consisting of "never," "some," "often," or "always."

Mogge and LePage (2004) state that the ADI was developed to be a brief and valid measure of depressive symptoms because common measures of depression either have no validity scales (e.g., BDI-II) or are lengthy (e.g., MMPI-2). In their initial validation studies on the ADI, using community samples instructed to fake depressive symptoms and psychiatric inpatient samples instructed to answer honestly, they found cut-scores of > 39 and > 49 on the depression scale for moderate and severe depression, respectively, have a .90 correlation with corresponding recommended cut-scores on the BDI-II. Its average inter-item correlation was .41. Also, using a cut-score of > 14, the feigning scale had a 79% sensitivity (the percentage of those instructed to feign depression identified by the ADI as feigning depression) rate while maintaining a 97% specificity (the percentage of individuals instructed to answer honestly identified by the ADI as answering honestly) rate. Based on more recent results in a psychiatric inpatient setting, Mogge has suggested a cut-score of > 13 may be even more effective (personal communication, July 6, 2006). While promising, these results require further validation.

Research Methods

There are three main designs employed when researching the validity of measures that evaluate malingering (Rogers & Bender, 2003). Differential prevalence design simply compares the responses between groups that are assumed to be different. For example, individuals who are seeking financial compensation are assumed to be more motivated to malinger than individuals in a similar condition not seeking financial compensation. This design is not methodologically sound, however, because it does not identify the proportion of feigning in each group and performance differences between groups.

Simulation, or analogue design, research is where participants are randomly assigned to honest or feigning conditions. This is the most common design used, primarily because of the relative ease of recruiting participants (typically undergraduates are offered extra credit), but also because of the strong internal validity that comes from establishing clear-cut differences between groups. The major shortcoming of these studies, however, is their very limited external validity. For instance, even though many of these studies provide small incentives (e.g., $5) for participants who feign symptoms in a believable manner, their motivation to perform well is likely to be much lower than someone who has the prospect of receiving enough remuneration that employment will no longer be a necessity.

Known-groups comparison research uses independent classification by mental health professionals to categorize malingerers and genuine patients. The strengths and weaknesses of this design are basically the opposite of the simulation design. By studying actual or suspected malingerers in real-world settings these studies capitalize on external validity. However, short of a confession, there is no "gold standard" criterion to establish malingerers from non-malingerers. Therefore, the internal validity of these studies is limited because every method used to differentiate honest from feigned symptom presentation is likely to be inaccurate to some extent, so there is always the possibility that some participants in the honest condition could actually be malingering and vice versa.

If a measure of malingering is valid there should be a convergence of results across both simulation and known-groups designs. Therefore, a study that incorporates both of these is the best option for addressing internal and external validity (Rogers, & Cruise, 1998).

Rationale

A two-stage approach to malingering may be a sufficiently sensitive, yet specific, way to assess malingering among disability seekers in a practical amount of time. In the first stage clients are screened for

malingering using a screening measure that is very sensitive with high NPP. In other words, a very high percentage of people who are classified as non-feigning or honest (as a result of scoring below a particular cut-score) are, indeed, honest. Therefore, those who fall below this cut-score on the screening measure are presumed not to be feigning and receive no further testing in that regard.

However, because of the high sensitivity of the screener, there may be many non-feigners who score above this cut-score. Therefore, those who are above the cut-score are considered indeterminate pending their results on a more in-depth method of testing that has high PPP. In other words, those classified as feigning (by scoring above a particular cut score) truly are feigning. A screening measure would prove especially useful if a cut-score could be established that has very high NPP while still substantially reducing the number of individuals who need further testing.

A two-stage strategy like this could significantly save clinical resources, but still maintain both high NPP and high PPP in the detection of malingering. As described earlier, Lewis et al. (2003) demonstrated the potential utility of the SIMS in this approach in detecting malingering among individuals who were being evaluated for competency or criminal responsibility. Also, Jackson, Rogers, and Sewell (2005) used this same design in examining the utility of the M-FAST (Miller Forensic Assessment of Symptoms Test; Miller, 2001) as a screen for detecting malingering among individuals being evaluated for competency to stand trial.

Using both a simulation and known-groups design, the purpose of the present study was to examine the utility of the SIMS and the ADI as malingering screening assessments among disability seekers. Currently neither of these measures have been validated with this population. Additionally, the only published validation studies on the ADI have been by its authors.

<div align="center">Method</div>

Participants

Clinical Component. Individuals who claimed disability based on functional impairment, and not MR (mental retardation) or TBI (traumatic brain injury), were administered the SIRS, SIMS, and ADI at a private psychological practice as part of a social security disability evaluation. The data from sixty-four individuals evaluated in this manner were used in this study. This sample consisted of 23 males (38%) and 41 females (62%) with an average age of 42.54 ($SD = 11.88$) and 11.9 ($SD = .59$) years of education. Twenty-one (32.8%) of these individuals had never been married, 18 (28.1%) were currently married, and 25 (39.1%) were divorced or separated.

With the exception of one African American, they were all Caucasian. For fifty-four (84.4%) of these individuals the primary diagnosis assigned by the psychologist after the standard disability evaluation was a mood disorder (Major Depressive Disorder, 34; Bipolar Disorder, 11; Dysthymia, 7; Adjustment Disorder w/ Mood, 1; and Mood Disorder NOS, 1). The remainder, with the exception of one Attention-Deficit/Hyperactivity Disorder (adult) diagnosis, had an anxiety disorder as their primary diagnosis (Anxiety Disorder NOS, 4; Panic Disorder, 2; Post-Traumatic Stress Disorder, 2; and Social Anxiety, 1).

Simulation Component. Sixty-three participants from the community were offered $5 for voluntary participation. This sample-of-convenience was drawn from workplaces and church groups recommended by acquaintances of the primary researcher. This sample consisted of 20 males (32%) and 43 females (68%) with an average age of 39.93 (*SD* =15.18). Eighteen (28.6%) of the participants indicated they had never been married, 35 (55.5%) were currently married, and 10 (15.9%) were divorced or separated. Education level of the participants varied with 15 (23.8%) having post-baccalaureate training, 14 (22.2%) having a college degree, 18 (28.6%) having some college attendance, 14 (22.2%) having a high school diploma or equivalent, and 2 (3.2%) having less than a high school education. All participants in this component were Caucasian.

Measures

Description of the SIRS. The Structured Interview of Reported Symptoms (SIRS) was developed to assess systematically deliberate distortions in the self-report of symptoms (Rogers et al., 1992). Although, it also has scales to assess response styles such as defensiveness and inconsistency, its emphasis is on feigning. It is a 45-60 minute structured interview of 172 items designed to be administered by a trained clinician. Results from detailed and general inquiries are scored on eight primary scales, each using a specific feigning detection strategy.

The first scale, RS, detects feigning by noting an endorsement of symptoms that are very infrequently endorsed by clinical populations (e.g., Does your face look strange or unfamiliar to you?). The SC scale measures how often the client endorses two symptoms that commonly occur alone, but are rarely paired together (e.g., At times when you feel hopeless, do you also feel lightheaded or faint). Endorsement of improbable or absurd symptoms are measured on the IA scale (e.g., Does pain ever bounce through your body going back and forth from limb to limb?). Malingerers also have a tendency to endorse a high proportion of obvious or blatant symptoms, which are measured on the BL scale (e.g., Do you have any major difficulties with having thoughts about suicide?). Conversely, the SU scale measures the under-reporting of common yet subtle symptoms (e.g., Do you have any

major difficulties with making everyday decisions?). The SEL scale is used to determine whether the client is responding to queries about blatant and subtle symptoms in a selective (versus indiscriminate) manner. The SEV scale measures if an inordinate number of blatant and subtle symptoms are reported as severe or unbearable. The last primary scale, RO, measures discrepancies between the client's reported symptoms and the behavioral observations of the clinician (e.g., Do you find it difficult to sit in a chair without looking under or behind it?).

Results from each primary scale are classified into one of four categories based on normative data: honest, indeterminate, probable, and definite. Overall results are classified as "feigning" if three or more primary scales are in the probable range or one or more in the definite range, "honest" if six or more scales are in the honest range, and "indeterminate" if none of these requirements are met. If a primary scale configuration is indeterminate, a total score can also be used. Scores above 76 are classified as feigning, scores below 73 are classified as honest, while scores in between remain indeterminate. This recommended criteria for the identification of feigning on the SIRS has a PPP of .98 (Rogers, 1992).

The SIRS has been established as a highly reliable measure that is cross-validated on diverse populations using known-groups and simulation designs in a variety of contexts (Rogers, 2001). The mean alpha coefficients are .86 on the primary scales and it has a mean interrater reliability of .96 (Rogers et al., 1992). An explicit objective of the SIRS is to minimize the number of false positives because of the potentially devastating consequences of miscategorizing a genuine patient as a feigner. Therefore, positive predictive power is of utmost importance.

Description of the SIMS. The Structured Inventory of Malingered Symptomatology (SIMS; Widows & Smith, 2005) is a self-report screening instrument for malingering. It is a 75-item true-false test that can be administered in about 15 minutes. The SIMS has five subscales: Affective Disorders (AF), Amnestic Disorders (AM), Low Intelligence (LI), Neurologic Impairment (N), and Psychosis (P). Each subscale contains 15 items designed to identify malingering of that particular type of disability (e.g., "Walking is difficult because of my problems with balance" = NI, "I seldom laugh" = AF, "People can put thoughts in my mind against my will" = P, "The United States has 55 states" = LI, "I can remember what I was doing one hour ago" = AM). A total score is also calculated as the sum of raw scores from all five scales. A total score of > 14 is suggested by the SIMS authors (Widows and Smith) to be indicative of possible malingering.

Description of the ADI. The Assessment of Depression Inventory is a 39-item, self-report measure that takes about five minutes to complete (Mogge & LePage, 2004). It assesses the presence of real depressive

symptomology and contains validity scales for feigned, random, and unreliable responding. Respondents are asked how frequently in the last two weeks they have experienced the feelings and reactions defined in the 39 items and they rate each item on a four-point scale consisting of "never," "some," "often," or "always" (i.e., never = 1, some = 2, often = 3, and always = 4). Some example items are: "I feel empty" and "I have problems with sleep" for depression, "I am so depressed I forget my name" and "I feel as if I and all my family deserve to die" for feigning, and "I enjoy paying taxes" and "Monday follows Sunday" for random responding. The reliability scale measures the discrepancy between responses to six pairs of questions that should be highly correlated. Unlike the SIMS, there is not a "total" score. Rather, there are four separate scores: one for depression and one for each of the three validity scales. Depression scores above 39 and 49 are considered to be indicative of moderate and severe depression, respectively. Feigning is suggested if the score on that scale is > 13. Cut-scores of > 7 and > 8 have been suggested for the random and reliability scales, respectively.

Description of the SCL-90-R. The Symptom Checklist-90-Revised is a 90-item self-report measure that assesses emotional functioning and psychopathology (Derogatis, 1975). Similar to the ADI, examinees are asked to report on a four-point scale (0 = not at all, 4 = extremely) how much they were distressed by a variety of symptoms in the previous week. It has nine clinical scales (Somatization, Obsessive-Compulsive, Interpersonal Sensitivity, Depression, Anxiety, Hostility, Phobic Anxiety, Paranoid Ideation, Psychotocism) and three global indices (Global Severity Index, Positive Symptom Distress Index, and Positive Symptom Total). It is best considered as a global measure of emotional distress because there is considerable overlap among scales (Edens, Otto, & Dwyer, 1999).

Procedure

Prior to recruiting participants for the simulation component, Institutional Review Board (IRB) approval was received from West Virginia University. For the clinical component, IRB approval was also received from West Virginia University, to use archival data from a local psychologist's private practice.

Clinical Component. During the period of IRB approval, the SIRS, SIMS, and ADI were routinely administered in Social Security disability evaluations. These measures were typically administered in the order listed above following a standard disability evaluation interview with a psychologist. For this study, individual results were assigned to an honest group or to a suspected malingering group based on SIRS scores (i.e., six primary scales in the honest range or a total score < 73 = honest; one primary scale in the definite range, or three primary scales in

the probable range, or a total score > 76 = suspected malingering; all other primary scale configurations or scores = indeterminate).

Simulation Component. Participants in the simulation component met with the primary researcher individually or in small groups at their workplace, church, or home. First, all participants were presented with an informed consent letter and a HIPPA waiver of PHI information. They were then instructed to complete the SCL-90-R in order to screen participants who were currently experiencing clinically significant levels of mental distress. A random number table, with an equal proportion of odd and even numbers, was used to assign participants to the groups. If the next number on the list was even the participant was instructed to complete the SIMS and the ADI honestly. If the number was odd, the participant was given these instructions:

> Put yourself in the following situation: You are applying for social security disability. You do not have any serious problems, but you are to pretend to suffer from major depression which prevents you from being able to work. The results of these tests will help determine whether you will receive thousands of dollars in disability benefits. Remember, just presenting symptoms of depression is not enough; it has to greatly affect your ability to work. Be careful, however, there may be some trick questions designed to catch persons trying to fake.

For additional incentive, participants in this condition were told that if they "successfully" malingered (i.e., obtained scores indicative of severe depression, but not indicative of malingering) they would be put in a drawing to win $50. Four participants met this criteria based on ADI scores (i.e., ADI Depression scale score > 49 and an ADI Feigning scale score < 14). The names of these four participants were placed in a drawing and the winner was paid $50.

After completing the SIMS and the ADI all participants were asked to fill-out a brief demographic questionnaire (Appendix A) that also included items to assess if they remembered and followed the instructions they were given.

<div align="center">Results</div>

Preliminary Analyses

Clinical Component. Using the SIRS scoring procedure described previously, 36 individuals were classified as honest and 20 were classified as suspected malingerers. The remaining eight individuals had indeterminate SIRS profiles and were excluded from further analyses.

Simulation Component. No participants were eliminated from the simulation sample for failure to

remember or follow directions (as indicated by their responses on the post-test questionnaire). However, three

participants (two who were instructed to answer honestly and one who was instructed to malinger) obtained

clinically elevated results on the SCL-90-R, which was defined as a GSI T-score > 62 or two dimensional T-scores

> 62. The data from these three participants were excluded from further analyses. As a result, 60 community

participants were retained for the present analyses, with 30 in the honest (control) group and 30 in the simulated

malingering group.

Discriminant Validity

Table 1 presents the means and standard deviations for each group on the SIMS and ADI scales.

Table 1

Means and Standard Deviations of SIMS and ADI Scale Scores by Comparison Groups

	Honest Groups				Malingering Groups				
	Control		Clinical		Simulation		Suspected		
	($n = 30$)		($n = 35$)		($n = 30$)		($n = 20$)		
	M	SD	M	SD	M	SD	M	SD	$F(3, 116)$
Total	7.73^a	2.94	17.13^b	6.60	26.57^c	14.64	30.15^c	9.13	30.87***
NI	1.27^a	1.55	4.22^c	2.65	5.63^c	4.56	5.95^c	3.09	12.73***
AF	4.40^a	1.52	6.55^b	1.84	9.60^c	1.65	9.65^c	1.60	64.77***
P	0.17^a	0.38	0.48^a	1.31	2.70^c	4.47	3.40^c	3.50	8.26***
LI	1.40^a	1.50	2.58^{ac}	2.39	2.80^{ac}	2.27	3.60^c	2.37	4.59**
AM	0.47^a	0.78	3.29^b	2.65	5.83^c	4.89	7.35^c	2.85	23.74***
ADI Fg	8.20^a	0.48	9.77^a	2.27	13.53^c	5.21	13.00^c	2.52	17.63***
ADI Dep	27.77^a	6.60	44.00^b	9.65	55.93^c	10.00	53.79^c	8.41	60.58***
ADI Rd	4.07^a	.25	4.81^b	1.44	4.70^{ab}	1.32	4.11^{ab}	.46	3.70*
ADI Rel	2.10^a	1.30	3.00^{ab}	1.47	3.17^{ab}	2.32	3.58^b	1.68	3.30*

Note. For SIMS subscales, NI = Neurologic Impairment, AF = Affective Disorders, P = Psychosis, LI = Low Intelligence, AM = Amnestic Disorders. For ADI scales: Fg = Feigning, Rd = Random, Rel = Reliable. Means in the same row that do not share superscripts differ in the Tukey honestly significant difference comparison. * = $p < .05$, ** = $p < .01$, *** = $p < .001$

A one-way analysis of variance (ANOVA) was conducted to examine group differences on the SIMS total score. A

significant effect was found, $F(3, 115) = 29.041$, $p < .001$. Post hoc analyses (Tukey honestly significant different

test) revealed that both the suspected malingering group ($M = 31.47$, $SD = 9.42$) and the simulated malingering

group ($M = 26.57$, $SD = 14.64$) scored significantly higher on the SIMS total than both the clinical honest group

($M = 17.77$, $SD = 8.92$) and the control group ($M = 7.57$, $SD = 2.93$), with all ps < .01. Thus, feigners endorsed

significantly more items on the SIMS than honest respondents. Interestingly, however, there was also a significant

difference between the clinical honest group and the control group ($p < .01$), but not between the two feigning
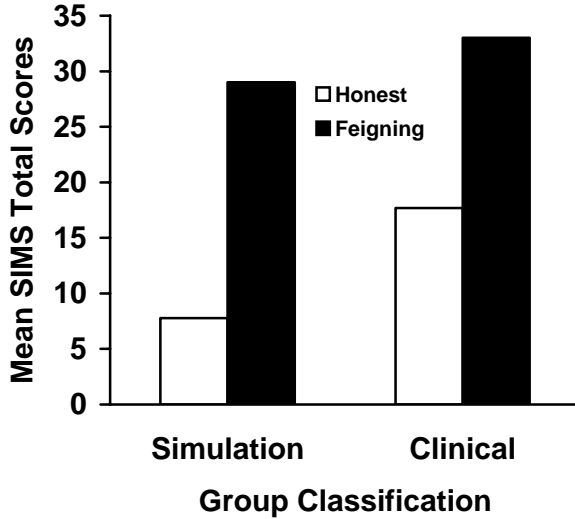
groups (see Figure 1).

Figure 1

One-way analyses of variance also revealed significant differences between mean group scores on all SIMS subscales: Fs(3,115) = 12.82, p < .001 (NI); 57.98, p < .001 (AF); 8.91, p < .001 (P); 5.85, p < .01 (LI); and 20.43, p < .001 (AM). Tukey's post hoc analyses demonstrated that both suspected and simulated malingering groups scored significantly higher than the control group on all subscales (ps < .001) except on LI where the simulated malingering group did not differ significantly. For subscales AF, P, and AM both feigning groups also scored significantly higher than the clinical honest group (ps < .05). However, for subscales NI and LI there were no significant differences between either feigning group and the clinical honest group. There were significant differences between the clinical honest group and the control group on subscales NI, AF, and AM (ps < .01), but not on subscales P and LI. There were no significant differences between the feigning groups on any subscale.

A one-way analysis of variance revealed significant differences between mean group scores on the ADI Feigning (Fg) scale as well, F(3, 115) = 16.71, p < .001. Tukey's post hoc analyses revealed that both the suspected malingering group (M = 12.93, SD = 2.49) and the simulated malingering group (M = 13.53, SD = 5.21) scored significantly higher than both the clinical honest group (M = 9.77, SD = 2.27) and the control group (M = 8.2, SD = .48), with all ps < .05. Thus, feigners endorsed significantly more items on the Fg scale than honest respondents, but neither the two honest groups nor the two feigning groups differed significantly from each other (see Figure 2).
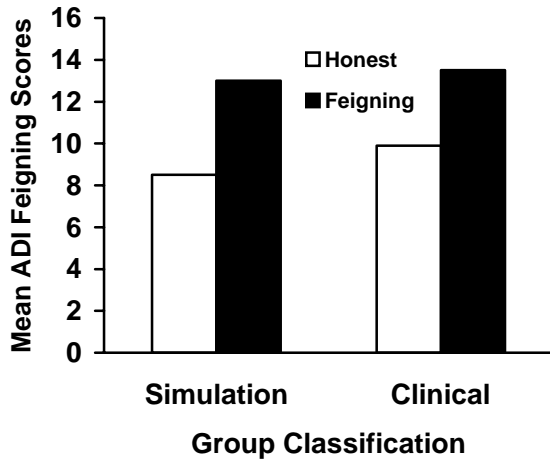
Figure 2

Table 1 also shows the mean scores for each group on the ADI Dep (depression) scale and the other two ADI validity scales, Rd (random responding) and Rel (unreliable responding). It should be noted, however, that these scales were not designed to detect feigning and the validity and reliability of these scales in measuring their respective construct (i.e., depression, random responding, and unreliable responding) was not examined in this study.

A one-way analysis of variance revealed significant group differences on the Dep scale, $F(3, 115) = 60.58$, $p < .001$. Tukey's post hoc analyses revealed that both the suspected malingering group ($M = 53.79$, $SD = 8.41$) and the simulated malingering group ($M = 55.93$, $SD = 10.00$) scored significantly higher on the Dep scale than both the clinical honest group ($M = 44.00$, $SD = 9.65$) and the control group ($M = 27.77$, $SD = 6.60$), with all ps $< .01$. Additionally, as might be expected, there was a significant difference between the clinical honest group and the control group ($p < .01$), but not between the feigning groups. One-way analyses of variance also revealed significant group differences on the Rd scale, $F(3, 115) = 3.70$, $p < .05$, and the Rel scale, $F(3, 115) = 3.30$, $p < .05$. Tukey's post hoc analyses revealed that for the Rd scale only the control group ($M = 4.07$, $SD = .25$) and the clinical honest group ($M = 4.81$, $SD = 1.44$) significantly differed, while for Rel scale only the control group ($M = 2.10$, $SD = 1.30$) and the suspected malingering group ($M = 3.58$, $SD = 1.68$) significantly differed.

As shown in Table 2, the SIMS Total and the ADI Feigning scales yielded large effect sizes (≥ .80) in the predicted direction for both feigning groups versus both honest groups. For feigners versus the control group, all SIMS subscales also yielded large effect sizes, except for subscale LI for simulated malingerers which had a medium effect size (≥ .50). For feigners versus the clinical honest group, on the other hand, only subscales AF, P, and AM for suspected malingerers yielded large effect sizes. Once again, subscale LI for simulated malingerers was the weak link, with all the other subscales yielding at least small (≥ .20) to medium effect sizes. Overall, the suspected malingering group (*Md* = 1.81) had higher effect sizes than the simulated malingering group versus honest respondents (*Md* = 1.16).

Table 2

Effect Sizes for SIMS and ADI Scale Scores Between Comparison Groups

| SIMS | Feigning vs. control | | Feigning vs. honest | | |
	SIM vs. CON	SUS vs. CON	SIM vs. HON	SUS vs. HON	Md
Total	1.78	3.31	.83	1.63	1.89
NI	1.28	1.91	.39	.60	1.05
AF	3.28	3.36	1.75	1.80	2.55
P	.80	1.30	.67	1.10	.97
LI	.73	1.11	.09	.43	.59
AM	1.53	3.37	.66	1.48	1.76
ADI Fg	1.44	2.65	.94	1.35	1.60
ADI Dep	3.32	3.44	1.21	1.08	2.26
ADI Rd	.66	.11	.08	.65	.38
ADI Rel	.57	.99	.09	.37	.51

Note. CON = control, HON = clinical honest, SIM = simulated malingering, SUS = suspected malingering. For SIMS subscales: NI = Neurologic Impariment, AF = Affective Disorders, P = Psychosis, LI = Low Intelligence, AM = Amnestic Disorders. For ADI scales: Fg = Feigning, Dep = Depression, Rd = Random, Rel = Reliability.

Utility Estimates

The purpose of this study was to examine the usefulness of the SIMS and the ADI as screening instruments for malingering within the context of a disability evaluation. Recalculating utility estimates in various contexts is useful because of potential changes in base rates (Jackson et al., 2005). Tables 3 – 8 display the sensitivity, specificity, positive predictive power (PPP) and negative predictive power (NPP), hit rate and area under the curve (AUC; Receiver Operator Characteristics) of the current recommended SIMS total and ADI Feigning scale cut-scores with alternative cut-scores.

Table 3

Utility Estimates SIMS Total Cut Scores in the Clinical Sample

Cut Score	PPP	NPP	Sensitivity	Specificity	Hit rate	AUC
> 10	.40	1.00	1.00	.14	.45	.58
> 11	.42	1.00	1.00	.20	.46	.61
> 12	.43	1.00	1.00	.26	.52	.65
> 13	.45	1.00	1.00	.31	.55	.66
> 14	.48	1.00	1.00	.37	.60	.71
> 15	.54	1.00	1.00	.51	.69	.73
> 16	.53	.95	.95	.51	.67	.70
> 17	.58	.95	.95	.60	.73	.74
> 18	.58	.92	.90	.63	.73	.72
> 19	.62	.92	.90	.69	.76	.74
> 20	.63	.89	.85	.71	.76	.73

Note. PPP = Positive Predictive Power, NPP = Negative Predictive Power, AUC = Area Under the Curve (Receiver Operator Characteristics)

Table 4

Utility Estimates SIMS Total Cut Scores in the Simulation Sample

Cut Score	PPP	NPP	Sensitivity	Specificity	Hit rate	AUC
> 10	.84	.89	.90	.83	.87	.87
> 11	.90	.90	.90	.90	.90	.90
> 12	.96	.88	.87	.87	.92	.92
> 13	.96	.83	.80	.97	.88	.90
> 14	.96	.81	.77	.97	.87	.87
> 15	1.00	.77	.70	1.00	.85	.87
> 16	1.00	.75	.67	1.00	.83	.83
> 17	1.00	.73	.63	1.00	.82	.82
> 18	1.00	.68	.53	1.00	.77	.77
> 19	1.00	.68	.53	1.00	.77	.77
> 20	1.00	.67	.53	1.00	.75	.75

Note. PPP = Positive Predictive Power, NPP = Negative Predictive Power, AUC = Area Under the Curve (Receiver Operator Characteristics)

Table 5

Utility Estimates for SIMS Total Cut Scores in the Total Sample

Cut Score	PPP	NPP	Sensitivity	Specificity	Hit rate	AUC
> 10	.57	.91	.94	.46	.66	.72
> 11	.60	.92	.94	.46	.70	.75
> 12	.63	.94	.92	.58	.73	.78
> 13	.64	.89	.90	.62	.74	.78
> 14	.65	.86	.86	.65	.74	.79
> 15	.71	.86	.84	.74	.78	.79
> 16	.70	.86	.83	.74	.76	.76
> 17	.73	.81	.76	.78	.77	.77
> 18	.72	.76	.68	.80	.75	.73
> 19	.76	.77	.68	.83	.76	.72
> 20	.78	.77	.62	.86	.76	.72

Note. PPP = Positive Predictive Power, NPP = Negative Predictive Power, AUC = Area Under the Curve (Receiver Operator Characteristics)

Table 6

Utility Estimates for ADI Feigning Scale Cut Scores in the Clinical Sample

Cut Score	PPP	NPP	Sensitivity	Specificity	Hit rate	AUC
> 8	.56	1.00	1.00	.54	.71	.69
> 9	.67	.93	.90	.74	.80	.77
> 10	.65	.84	.75	.77	.76	.72
> 11	.71	.85	.75	.83	.80	.75
> 12	.86	.80	.60	.94	.82	.75
> 13	.86	.80	.60	.94	.82	.75
> 14	.83	.69	.25	.97	.71	.61

Note. PPP = Positive Predictive Power, NPP = Negative Predictive Power, AUC = Area Under the Curve (Receiver Operator Characteristics)

Table 7

Utility Estimates for ADI Feigning Scale Cut Scores in the Simulation Sample

Cut Score	PPP	NPP	Sensitivity	Specificity	Hit rate	AUC
> 8	.84	.86	.87	.83	.85	.85
> 9	.95	.74	.67	.97	.82	.81
> 10	1.00	.71	.60	1.00	.80	.80
> 11	1.00	.70	.57	1.00	.78	.80
> 12	1.00	.67	.50	1.00	.75	.78
> 13	1.00	.67	.50	1.00	.75	.78
> 14	1.00	.63	.40	1.00	.70	.70

Note. PPP = Positive Predictive Power, NPP = Negative Predictive Power, AUC = Area Under the Curve (Receiver Operator Characteristics)

Table 8

Utility Estimates for ADI Feigning Scale Cut Scores in the Total Sample

Cut Score	PPP	NPP	Sensitivity	Specificity	Hit rate	AUC
> 8	.67	.92	.92	.68	.78	.77
> 9	.79	.82	.76	.85	.81	.79
> 10	.80	.77	.66	.88	.78	.76
> 11	.84	.77	.64	.91	.79	.76
> 12	.93	.73	.54	.97	.78	.75
> 13	.93	.73	.54	.97	.78	.75
> 14	.94	.66	.34	.98	.70	.67

Note. PPP = Positive Predictive Power, NPP = Negative Predictive Power, AUC = Area Under the Curve (Receiver Operator Characteristics)

For the clinical sample (see Table 3), the recommended cut-score on the SIMS Total (> 14) has perfect sensitivity (1.00), but only modest specificity (.37). Increasing the cut-score to > 19 greatly increases specificity (.69), while maintaining high sensitivity (.90). In the simulation sample (see Table 4), on the other hand, decreasing the cut-score (e.g., > 12) provides more optimal results. For the total sample (see Table 5), the Rogers et al. (1996) recommended cut-score of > 16 appears to be adequate with a sensitivity of .83 and a specificity of .74.

Inspection of Tables 6 and 7 show the ADI Feigning scale recommended cut-score of > 13 has excellent specificity for both clinical and simulation samples (.94 and 1.00, respectively), but less than ideal sensitivity (.60

and .50) for a screening measure. For both samples, separately and combined (see Table 8), a cut-score of > 9 appears to be more effective. Utility estimates cannot be made of lower cut-scores because a minimum score of 8 is obtained even when all scale items are endorsed as "never" (non-feigning).

The possibility of the SIMS and the ADI providing incremental validity when used in combination was also examined. As shown in Table 9, when using the most optimal cut-scores previously discussed (i.e., SIMS total > 19, ADI Feigning > 9), the SIMS and ADI combined (cut-score > 28) had hit rates equal or slightly higher than the SIMS alone. However, these hit rates were equal or slightly lower than the ADI alone.

Table 9

Hit Rate of the SIMS Total and ADI Feigning Scale Scores

Participant Condition	SIMS (> 19)	ADI (> 9)	SIMS + ADI (> 28)
Simulation Sample	.77	.82	.82
Clinical Sample	.76	.80	.76
Total Sample	.75	.81	.79

The utility estimates of the SIMS subscales (utilizing the recommended cut-scores) are presented in Tables 10-12. Overall, these results support the finding of previous research on the SIMS that suggest the SIMS Total score is a more effective screen than any particular subscale. Although, in the clinical sample, the P subscale had a hit rate (.84) superior to any SIMS Total cut-score, its relatively low sensitivity (.65) precludes it from being an effective screening measure.

Table 10

Utility Estimates for SIMS Subscale Cut Scores in the Clinical Sample

Scale/cut-score	PPP	NPP	Sensitivity	Specificity	Hit rate	AUC
NI > 2	.43	.78	.80	.40	.55	.56
AF > 5	.44	1.00	1.00	.29	.55	.60
P > 1	.87	.83	.65	.94	.84	.77
LI > 2	.52	.79	.70	.63	.65	.64
AM > 2	.56	1.00	1.00	.54	.71	.74

Note. NI = Neurologic Impairment, AF = Affective Disorders, P = Psychosis, LI = Low Intelligence, AM = Amnestic Disorders, PPP = Positive Predictive Power, NPP = Negative Predictive Power, AUC = Area Under the Curve (Receiver Operator Characteristics)

Table 11

Utility Estimates for SIMS Subscale Cut Scores in the Simulation Sample

Scale/cut score	PPP	NPP	Sensitivity	Specificity	Hit rate	AUC
NI > 2	.80	.71	.67	.83	.75	.75
AF > 5	.78	.96	.97	.73	.85	.85
P > 1	1.00	.60	.33	1.00	.67	.67
LI > 2	.68	.61	.50	.77	.63	.63
AM > 2	.94	.69	.57	.97	.77	.77

Note. NI = Neurologic Impairment, AF = Affective Disorders, P = Psychosis, LI = Low Intelligence, AM = Amnestic Disorders, PPP = Positive Predictive Power, NPP = Negative Predictive Power, AUC = Area Under the Curve (Receiver Operator Characteristics)

Table 12

Utility Estimates for SIMS Subscale Cut Scores in the Total Sample

Scale/cut score	PPP	NPP	Sensitivity	Specificity	Hit rate	AUC
NI > 2	.58	.74	.72	.60	.65	.65
AF > 5	.60	.97	.98	.49	.70	.72
P > 1	.92	.70	.46	.97	.75	.70
LI > 2	.59	.68	.58	.69	.64	.63
AM > 2	.69	.79	.74	.74	.74	.73

Note. NI = Neurologic Impairment, AF = Affective Disorders, P = Psychosis, LI = Low Intelligence, AM = Amnestic Disorders, PPP = Positive Predictive Power, NPP = Negative Predictive Power, AUC = Area Under the Curve (Receiver Operator Characteristics)

Table 13

Area Under the Curve for SIMS Subscale Combinations in the Clinical Sample

Subscales	Cut score	AUC	Subscales	Cut score	AUC
NI + AF	> 4	.61	NI + P + LI	> 5	.60
NI + P	> 3	.63	NI + P + AM	> 5	.63
N + LI	> 4	.62	NI + LI + AM	> 6	.58
NI + AM	> 4	.63	AF + P + LI	> 5	.71
AF + P	> 6	.74	AF + P + AM	> 8	.68
AF + LI	> 7	.68	AF + LI + AM	> 9	.65
AF + AM	> 7	.65	P + LI + AM	> 5	.73
P + LI	> 3	.79	NI + AF + P + LI	> 10	.61
P + AM	> 3	.74	NI + AF + P + AM	> 10	.63
LI + AM	> 4	.66	NI + AF + LI + AM	> 11	.61
NI + AF + P	> 8	.65	NI + P + LI + AM	> 11	.63
NI + AF + LI	> 9	.63	AF + P + LI + AM	> 10	.68
NI + AF + AM	> 9	.60	.		

Note. NI = Neurologic Impairment, AF = Affective Disorders, P = Psychosis, LI = Low Intelligence, AM = Amnestic Disorders, PPP = Positive Predictive Power, NPP = Negative Predictive Power, AUC = Area Under the Curve (Receiver Operator Characteristics)

Lastly, the Area Under the Curve (AUC) for all possible SIMS subscale combinations (using the sum of the inclusive subscale cut-scores) was calculated in the clinical sample to determine if removing one or more of the subscales led to increased utility rates. As can be seen in Table 13, only the combination of subscales P and LI yielded an AUC (.79) superior to what is obtained when using a cut-score of > 19 on the SIMS Total (.76). Further analysis revealed that this combination had .67 PPP, .93 NPP, .90 sensitivity, .74 specificity, and .80 hit rate. All of

these results are either equal or slightly superior to what is obtained with a SIMS Total cut-score of > 19 (.62 PPP, .92 NPP, .90 sensitivity, .69 specificity, and .76 hit rate).

Discussion

Overall results suggest the SIMS and the ADI can help discriminate between honest and malingering disability seekers. Significant differences were found between the feigning and the honest groups on the SIMS total score and the ADI Feigning scale score in both the clinical and simulation samples. Specifically, feigners tend to endorse more symptoms and obtain higher scores on both of these scales than honest respondents.

The utility estimates of various cut-scores on these scales were examined for the simulation, clinical, and total samples. However, the results in the clinical sample alone are arguably the most relevant because they have greatest external validity. Therefore, increasing the SIMS total cut-score to > 19 appears to provide more optimal results in a clinical setting than previous recommended cut-scores (> 14 or > 16). Conversely, decreasing the ADI Feigning cut-score to > 9 may be more effective than the current recommended cut-score (> 13).

Both suspected and simulated malingerers also scored significantly higher than controls on all five SIMS subscales (except simulated malingerers on the LI [Low Intelligence] scale). Although simulated malingerers were specifically instructed to feign depression and the majority of suspected malingerers were diagnosed with a mood disorder, the AF (Affective Disorders) subscale was only slightly more effective than the SIMS total score in differentiating feigners from non-feigners in the community sample and less effective than the SIMS total in the clinical sample. Consistent with the findings of previous researchers (e.g., Poythress et al., 2001) the results of this study suggest the SIMS subscales in isolation or in various combinations are less effective than the SIMS total score in differentiating feigners from non-feigners. The only possible exception to this appears to be the combination of P (Psychosis) and LI (Low Intelligence) which had utility rates slightly superior to the SIMS total score in the clinical sample.

In addition, as expected, there were not significant differences between the simulation and suspected malingering groups on either the SIMS Total or the ADI Feigning scale. These results suggest that the simulated malingerers answered the SIMS and ADI items in a manner similar to an actual disability malingerer and that the suspected malingerers answered in a manner similar to a person deliberately feigning for monetary incentive. Thus, providing convergent validity.

Interestingly, however, there were significant differences between the clinical honest group and the control group on these same scales. The community participants endorsed less items and obtained lower SIMS total and ADI Feigning scales scores than honest (based on SIRS scores) disability seekers. The fact that all community participants were free of clinical levels of mental illness or distress (according to SCL-90-R scores) and none of the clinical participants were (based on an Axis I diagnosis assigned by a psychologist) seem to suggest that genuine mental illness leads to elevated scores on both of these measures. This is somewhat surprising because neither of these scales were designed to be sensitive to genuine psychopathology. However, these results are consistent with Edens et al. (1999) who also found a correlation between bona fide distress and raised SIMS scores. The difference among these groups is especially striking when it is noted that only 3.3% of the honest community participants (one person out of 30) obtained a SIMS total score > 14 (this individual scored a 15) while 68.8% of the honest clinical participants (22 out of 35 people) obtained a SIMS total score > 14. Because of its overall lower sensitivity, the difference in false-positives was not as remarkable for the ADI Feigning scale (cut-score > 13) with none (0 out of 30) among simulation respondents and 5.7% (2 out 35) among the clinical respondents.

Another possible explanation for these results is that disability seekers, even those who have genuine mental illness and are generally honest, tend to exaggerate their symptoms to ensure they will be considered functionally impaired by the disability evaluator. This is similar to the conclusion reached by Geraerts et al. (2006) when they found that individuals who report memories (recovered or continuous) of childhood sexual abuse (CSA) score significantly higher on the SIMS than controls. They reasoned that elevated SIMS scores were more likely indicative of CSA respondents trying to make a convincing impression on the researchers rather than genuine mental illness. However, this explanation seems to make more sense in the present study because in the Geraerts et al. study there were no external incentives for feigning.

The external validity of simulation or analogue research is always questionable. Although participants were told they would be put in a drawing to win $50, this sum is likely to be perceived as an insignificant amount of money and/or the likelihood of winning may have seemed remote. Thus, it seems plausible that the motivation of the simulated malingerers to feign successfully may not have approximated that of the suspected malingerers.

For the clinical sample it is the internal validity, or the accuracy of the SIRS in discriminating between honest and feigning disability seekers, that is uncertain. Especially considering the SIRS has primarily been validated in identifying feigned psychosis. Although some of the disability seekers reported some symptoms of

psychosis, the primary diagnosis for all of them was a mood or anxiety disorder (except one ADHD). Therefore, the SIRS may not be the "gold standard" among this population. Interestingly, however, this would suggest the SIRS may be too specific. Thus, the 36% rate of suspected malingering in this sample may actually be a conservative estimate.

Both concerns of internal and external validity are attenuated to an extent because there were no significant findings between the simulated and malingering groups on any SIMS scale or on the ADI Feigning scale. Additionally, the 36% malingering rate is consistent with the findings of Mittenberg et al. (2002), and Miller and Davilla (2001) who reported malingering rates of 30% and 37.5%, respectively, among disability seekers.

Future research that has a control group more analogous to the clinical group in terms of mental illness may help clarify the issue of whether genuine mental illness can lead to SIMS scores above the feigning cut-score or if "honest" disability seekers have a tendency to feign. For instance, instead of excluding individuals who have depression or anxiety from the control group, exclude individuals who do *not* have depression or anxiety. The control group scores would then provide strong evidence for or against the idea that mental illness can lead to a SIMS total score above the cut-score.

The most promising direction for future research on malingered psychological disability is with known-groups of actual or suspected malingerers, but internal validity can be dubious. The development of an extensive "gold standard" assessment of mood or anxiety disorders may be necessary to sufficiently address the issue of internal validity of studies involving actual or suspected malingering disability seekers.

Another potential limitation of this study that could be addressed in future research is its generalizability to minority populations because both the clinical and simulation samples (with the exception of one African-American in the clinical sample) were exclusively Caucasian.

The primary purpose of this study was to examine the utility of the SIMS and the ADI in clinical use. Optimal cut-scores on a screening instrument vary according to context. The relative importance of classificatory accuracy must be weighed against practical constraints (e.g., time, money). The original authors' previous recommended SIMS and ADI cut-scores (> 14 and > 13, respectively) may be effective for certain populations and settings (e.g., forensic, inpatient), but it seems clear that they need to be adjusted for a disability-seeking population. For the SIMS, using the current recommended cut-score of > 14, or even > 16 (as suggested by Rogers et al., 1996), results in too many false-positives to be considered practical for routine clinician use. Conversely, the ADI Feigning

scale cut-score of > 13 results in too many false-negatives to be an effective screening measure. In the present study, increasing the SIMS total cut-score to > 19 and decreasing the ADI Feigning scale cut-score to > 9 yields more optimal results.

Ideally a screening instrument cut-score would yield nearly perfect sensitivity and NPP, but it must have reasonable specificity and PPP to be useful. To look these results another way, using a SIMS total cut-score of > 14 with the actual disability seekers in this study would have excluded only 20%, or 11 out of 55, of them from requiring more extensive malingering testing (e.g., SIRS). Administering 55 SIMS might be slightly more time efficient than administering 11 SIRS, but not particularly expedient. However, using a cut-score of > 19 with this sample would have eliminated nearly half (47.3%) of these individuals from needing more extensive testing, and would have resulted in only two false-negatives.

Lastly, it should be emphasized again that the SIMS and the ADI are not intended to be diagnostic. Positive results on these measures should be followed-up with more extensive testing that has higher specificity and PPP before feigning is suspected. Also, psychological testing alone cannot provide a definitive diagnosis of malingering. A multi-method approach that also integrates information obtained from interviews, behavioral observations, and collateral reports must be used to confirm the likelihood of feigning and to identify an external incentive as the motivation for the feigning (Rogers, 1996). Furthermore, scoring high on a malingering assessment does not necessarily preclude genuine mental or medical illness that is functionally disabling. Therefore, a person who exaggerates or fabricates some psychological symptoms may still be rightfully entitled to disability benefits.

References

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

Derogatis, L.R. (1975). *Symptom Checklist-90 Revised* (SCL-90-R) *administration, scoring, and procedures manual*. Minneapolis, MN: NCS/Pearson

Edens, J.F., Otto, R.K., & Dwyer, T. (1999). Utility of the Structured Inventory of Malingered Symptomatology in identifying persons motivated to malinger Psychopathology. *Journal of the American Academy of Psychiatry and the Law, 27*(3), 387-396.

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

Heinze, M.C., & Purisch, A.D. (2001). Beneath the mask: Use of psychological tests to detect and subtype malingering in criminal defendants. *Journal of Forensic Psychology Practice, 1*(4), 23-52.

Jackson, R.L., Rogers, R., & Sewell, K.W. (2005). Forensic applications of the Miller Forensic Assessment of Symptoms Test (MFAST): Screening for feigned disorders in competency to stand trial evaluations. *Law and Human Behavior, 29*(2), 199-210.

Jelicic, M., Hessels, A., & Merckelbach, H. (2006). Detection of feigned psychosis with the structured inventory of malingered symptomatology (SIMS): A study of coached and uncoached simulators. *Journal of Psychopathology and Behavioral Assessment*, *28*, 19-22.

Lewis, J.L., Simcox, A.M., & Berry, D.T.R. (2002). Screening for feigned psychiatric symptoms in a forensic sample by using the MMPI-2 and the Structured Inventory of Malingered Symptomatology. *Psychological Assessment, 14,* 170-176.

Merckelbach, H., & Smith, G.P. (2003). Diagnostic accuracy of the Structured Inventory of Malingered Symptomatology (SIMS) in detecting instructed malingering. *Archives of Clinical Neuropsychology, 18*(2), 145-152.

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

Miller, H.A., & Davila, G.M. (2001). Preliminary analysis of the MFAST: Detecting malingering with disability claimants. Poster presented at the 47th annual Convention of the Southwestern Psychological Association.

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.

Mogge, N.L., & LePage, J.P. (2004). The assessment of depression inventory (ADI): A new instrument used to measure depression and to detect honesty of response. *Depression and Anxiety, 20,* 107-113.

Poythress, N.G., Edens, J.F., & Watkins, M.M. (2001). The relationship between psychopathic personality features and malingering symptoms of major mental illness. *Law and Human Behavior, 25*(6), 567-582.

Rogers, R. (2001). *Handbook of Diagnostic and Structured Interviewing.* New York: Guilford Press.

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

Rogers, R., & Bender, S. (2003). Evaluation of malingering and deception. In A.M. Goldstein & I. B. Weiner (Eds.), *Handbook of psychology* (pp. 109-129). New York: Wiley & Sons.

Rogers, R., & Cruise, K. R. (1998). Assessment of malingering with simulation designs: Threats to external validity. *Law and Human Behavior, 22*, 273-285.

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., Jackson, R.L., & Kaminski, P.L. (2005). Factitious psychological disorders: The overlooked response style in forensic evaluations. *Journal of Forensic Psychology Practice, 5*(1), 21-41.

Rogers, R., Kropp, P. R., Bagby, R.M., & Dickens, S.E. (1992). Faking specific disorders: A study of the Structured Interview of Reported Symptoms (SIRS). *Journal of Clinical Psychology, 48,* 643-648.

Rogers, R., Sewell, K.W., & Salekin, R. (1994). A meta-analysis of malingering on the MMPI-2, *Assessment, 1,* 227-237.

Smith, G.P., & Burger, G.K. (1997). Detection of malingering: A validation of the Structured Inventory of Malingered Symptomatology (SIMS). *Journal of the American Academy of Psychiatry and the Law, 25*(2), 183-789.

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

Appendix A

Participant Questionnaire

1. Age: _____

2. Gender: A. Male
                     B. Female

3. Race: _____

4. Marital Status: A. Never Married
                                    B. Married
                                    C. Divorced/Separated/Widowed

5. Education Level: A. did not finish high school
                                    B. high school graduate or equivalent
                                    C. some college
                                    D. college graduate
                                    E. graduate school/professional degree

6. Have you ever sought help from a mental health professional?
          A. Yes
          B. No
If yes, for what reason?
_____

7. Do you currently take any prescription medications for psychological reasons?
          A. Yes
          B. No
If yes, what medications?_____

8. In what way were you instructed to respond to the assessments in the current study?
          A. Honestly
          B. As if pretending that I suffer from major depression

9. In what way did you **actually respond** to the assessments in the current study (in other words, did you abide by the instructions)?
          A. Honestly
          B. As if pretending that I suffer from major depression

Carl B. Clegg

Curriculum Vitae

Address: 688 Powell Ave. D9
                      Morgantown, WV 26505
Telephone: (801) 216-5990
E-mail: cleggcarl@hotmail.com
Date of Birth: August 4, 1977
Place of Birth: Salt Lake City, Utah
Citizenship: United States of America

Education:
Utah Valley State College. Associate in Science. May 2002
University of Utah. Bachelor of Science. Psychology. May 2004

Academic Honors:
High Honors. Dean's List 2001-2002. Utah Valley State College.
Summa Cum Laude. Dean's List 2003-2004. University of Utah.
Psi Chi (National Honor Society in Psychology)

Professional Positions:
Clinical Practicum. William R. Sharpe, Jr. Hospital (Forensic Unit), Weston, WV. July 2006 to June 2007. Duties: administer psychological assessments and provide individual and group therapy to individuals adjudicated as incompetent to stand trial or not guilty by reason of mental illness. Supervisor: Neil Mogge, Ph.D.

Graduate Teaching Assistant. West Virginia University. August 2005 to May 2006. Duties: teach Psychology 101.

Plethysmographer, Center For Family Development. Salt Lake City, Utah. October 2002 to August 2005. Duties: administer written and physiological testing needed for psychological evaluations of sex offenders in jails and treatment centers. Supervisor: Larry Fox, Ph.D.

Supplemental Instruction Leader, University of Utah. January 2004 to December 2004. Duties: group tutoring, write test review sheets, and conduct test review sessions for Psychology 101. Supervisor: David Dodd, Ph.D.

Research Assistant, University of Utah. August 2003 to June 2005. Duties: interview adolescents in substance abuse treatment centers, collect and code data. Supervisor: Paul Florsheim, Ph.D.

Research Assistant, University of Utah. August 2003 to December 2003. Duties: train rats to perform a variety of tasks, collect data. Raymond Kesner, Ph.D.

Presentations and Publications:
Shiozaki, T., Hiraoka, G., Florsheim, P., Heavin, S., Clegg, C., & Tiffany, S. (in press). Craving among polysubstance-using adolescents. *Journal of Child and Adolescent Substance Abuse.*

Clegg, C., Giron, A., Thomas, T., Mogge, N, & Fremouw, W. (2007). Screening malingering among disability seeking outpatients. Poster accepted for presentation at the Southeastern Psychological Association Convention in New Orleans, LA.

McCoy, K., Fremouw, W., & Clegg, C. (2007). Self injurious and choking game behaviors among college students. Poster accepted for presentation at the Southeastern Psychological Association Convention in New Orleans, LA.

McCoy, K., Fremouw, W., Tyner, E., Clegg, C., Johannson-Love, J., & Strunk, J. (2006). Criminal thinking styles and illegal behavior among college students: Validation of the PICTS. *Journal of Forensic Sciences*, *51*(5), 1174-1177.

Shiozaki, T., Florsheim, P., Tiffany, S., Heavin, S., Minen, S., & Clegg, C. (2004). *Adolescent Substance Abuse: Can Personality Predict Craving Experience?* Paper presented at meeting of the Rocky Mountain Psychological Association. Reno, NV.

References:
William Fremouw, Ph.D., Professor, Department of Psychology, West Virginia University. Morgantown, WV. (304) 293-2001.   wfremouw@wvu.edu

Neil Mogge, Ph. D., Professor, Behavioral Medicine and Psychiatry, West Virginia University. Morgantown, WV. (304) 269-1210 x323.   nmogge@hsc.wvu.edu

Paul Florsheim, Ph.D., Associate Professor, Department of Psychology, University of Utah. Salt Lake City, Utah. (801) 585-5676.   florsheim@psych.utah.edu