Law Hum Behav (2007) 31:249–260
DOI 10.1007/s10979-006-9062-8

ORIGINAL ARTICLE

An Evaluation of Malingering Screens with Competency to Stand Trial Patients: A Known-Groups Comparison

Michael J. Vitacco · Richard Rogers · Jason Gabel · Janice Munizza

Published online: 21 October 2006
© American Psychology-Law Society/Division 41 of the American Psychological Association 2006

Abstract The assessment of malingering is a fundamental component of forensic evaluations that should be considered with each referral. In systematizing the evaluation of malingering, one option is the standardized administration of screens as an initial step. The current study assessed the effectiveness of three common screening measures: the Miller Forensic Assessment of Symptoms Test (M-FAST; Miller, 2001), the Structured Inventory of Malingered Symptomatology (SIMS; Widows & Smith, 2004), and the Evaluation of Competency to Stand Trial-Revised Atypical Presentation Scale (ECST-R ATP; Rogers, Tillbrook, & Sewell, 2004). Using the Structured Interview of Reported Symptoms (SIRS) as the external criterion, 100 patients involved in competency to stand trial evaluations were categorized as either probable malingerers ($n = 21$) or nonmalingerers ($n = 79$). Each malingering scale produced robust effect sizes in this known-groups comparison. Results are discussed in relation to the comprehensive assessment of malingering within a forensic context.

Keywords Malingering · Competency to stand trial · Screens

A fundamental component of pretrial forensic evaluations is the assessment of malingering (Rogers & Bender, 2003). Given the high stakes associated with pretrial forensic evaluations (e.g., lengthy prison sentences), clinicians cannot assume the veracity of their patients' presentations. Within the forensic context, recent estimates of malingering range from 15.7% (Rogers, 1997) to 17.4% (Rogers, Salekin, Sewell, Goldstein, & Leonard, 1998). Regarding pretrial issues, competence to stand trial (CST) evaluations eclipse other forensic referrals in their frequency with annual estimates ranging from 50,000 (Skeem, Golding, Berge, & Cohn, 1998) to 60,000

M. J. Vitacco (✉) · J. Munizza
Mendota Mental Health Institute, 301 Troy Drive, Madison, WI 53704
e-mail: vitacmj@dhfs.state.wi.us

R. Rogers
University of North Texas, Texas

J. Gabel
University of Wisconsin, Wisconsin

per year (Bonnie & Grisso, 2000). When the number of CST evaluations are combined with base-rate estimates of malingering (Rogers, Jackson, Sewell, & Harrison, 2004), thousands of CST evaluations occur annually in which issues of malingering must be systematically evaluated.

The need for systematic evaluation is critical because seasoned clinicians often rely on their own individualistic perspectives in deciding when to assess for malingering (Rogers et al., 1998). Unfortunately, the overreliance on clinical judgment can be the source of misdiagnosis and inaccurate information given to the court. As such, the accurate assessment of malingering is vitally important. Misclassifications of malingering can have far-reaching effects on legal outcomes and can also preclude the provision of basic mental health services to offenders with legitimate mental health disorders.

Assessing malingering in clinical evaluations

The American Psychiatric Association (2000) provided screening indices in DSM-IV TR that were not rigorously tested. These indices include (1) referral for a medicolegal evaluation, (2) marked discrepancy between claimed stress and objective findings, (3) lack of cooperation during the diagnostic evaluation, and (4) the presence of Antisocial Personality Disorder. Available data suggest that these indices may produce unacceptably high false-positives in the neighborhood of 80% (Rogers, 1999; Rogers & Vitacco, 2002). Fortunately, standardized methods for the assessment of malingering have been developed that rely on empirically-based detection strategies (Rogers, Jackson, Sewell, & Salekin, 2005). These empirically-based strategies have been used in developing both comprehensive instruments and specialized screens, which are designed to discriminate systematically between genuine patients and those feigning mental disorders across major diagnostic categories[1] (Rogers, Sewell, Martin, & Vitacco, 2003). The next paragraphs will review findings from some of the more widely used malingering instruments in forensic practice.

The Structured Interview of Reported Symptoms (SIRS; Rogers, Bagby, & Dickens, 1992) has been used as a criterion measure in the assessment of malingering. The SIRS, a structured interview, has been extensively validated using both simulation and known-group designs (Luis, 2003; Rogers et al., 2005; Rogers, Sewell, Grandjean, & Vitacco, 2002). Through the use of a multifaceted scoring system, the SIRS minimizes false negatives and demonstrates very high positive predictive power (PPP) at particular score levels (.98; Rogers et al., 1992). In assessing malingering, the SIRS scales are highly effective at differentiating feigners from patients with genuine disorders with the average Cohen's *d* of 1.74 for SIRS primary scales (Rogers, 1997). Due to its excellent psychometric properties, the SIRS is often used by experts in the field of forensic psychology for assessing malingering (Lally, 2003).

A recent development in evaluating malingered presentation is the creation of screens that are used to conserve time and financial resources. Two general malingering screens are the Miller Forensic Assessment of Symptoms Test (M-FAST; Miller, 2001), and the Structured Inventory of Malingered Symptomatology (SIMS; Widows & Smith, 2004). In addition, Rogers et al. (2004) constructed specialized scales as part of the Evaluation of Competency to Stand Trial-Revised (ECST-R), namely the Atypical Presentation Scales (ATP). The ATP is designed specifically to

[1] Individuals feigning cognitive impairment require completely different detection strategies. For specific strategies, please refer to Rogers (1997) and Rogers and Bender (2003).

assess feigned incompetency. Validation studies for these three malingering screens are examined in subsequent paragraphs.

Like the SIRS, the M-FAST has been tested via analogue and known-groups designs. Guy and Miller (2004) tested the M-FAST in a sample of 50 prisoners who requested placement to the mental health unit. Using a cut-score ≥ 6, the M-FAST produced moderately good PPP (.78) and negative predictive power (NPP = .78). Malingerers, as classified by the SIRS, had significantly higher scores than other inmates on the M-FAST. More recently, Jackson, Rogers, and Sewell (2005) tested various cut-scores of the M-FAST with the SIRS as the criterion measure. She and her colleagues recommended a cut score ≥ 6 on the total M-FAST score, which yielded positive results with high NPP (.91) and moderate Sensitivity (.76).[2]

The SIMS, a self-report instrument, is also frequently used in malingering evaluations. A total SIMS score ≥ 14 is recommended as a cut score for suspected malingering (Widows & Smith, 2004). Originally, a study by Smith and Burger (1997) yielded good utility estimates in a SIMS simulation study, although they relied entirely on college students and lacked a clinical comparison sample. Edens, Otto, and Dwyer (1999) tested the SIMS with a simulation design. While identifying most simulators, the authors were concerned about the high number of false positives. In a known groups-comparison with 64 federal prisoners undergoing pretrial evaluations, Lewis, Simcox, and Berry (2002) found that SIMS produced exceptional sensitivity (1.00) and NPP (1.00) with the SIRS used to define the criterion groups (i.e., malingering vs. nonmalingering). Lewis et al. (2002) found large effect sizes (i.e., Cohen's *d*s) on the SIMS ranging from 1.1 to 3.0. When used with highly sophisticated simulators (i.e., doctoral students in psychology), Rogers, Jackson, and Kaminski (2005) found that the SIMS was not especially effective, although the Neurologic Impairment scale (Cohen's d = 1.74) and total score (Cohen's d = 1.45) evidenced large effective sizes when compared to the control condition. In general, the SIMS appears effective with simulators that lack sophisticated preparation.

The final screen utilized in the present research is the ECST-R ATP scale. The ECST-R is a second-generation structured interview designed to assist clinicians in conducting competency to stand trial evaluations. One of the unique features of the ECST-R is the inclusion of 28 items designed to assess for competency-specific feigned impairment. Research by Rogers et al. (2002) using the ATP scale with jail detainees found significantly higher scores on all ATP scores (*d* scores ranging from 0.73 to 1.36), when the SIRS was used as the criterion measure for the classification of malingering. With a cut-score of >5, the ECST-R ATP demonstrated excellent sensitivity (.86) and NPP (.94) with 96 jail detainees tested in a simulation condition and 56 patients undergoing competency evaluations (see Rogers et al., 2004). In summary, recent research (Rogers et al., 2002; 2004) has found that the ATP scale is a useful screen for malingering, especially in CST evaluations.

The current study was designed to test the effectiveness of three malingering screens (M-FAST, SIMS, and ECST-R ATP) used to evaluate criminal forensic populations. We focused on three related issues. First, we evaluated the homogeneity of these scales by examining their internal consistencies and item-scale correlations. Second, we evaluated the discriminant validity of each malingering scale via effect sizes. Finally, we sought to determine which measures were most effective (i.e., sensitivity and NPP) at retaining suspected referrals for further evaluation.

[2] Jackson et al. (2004) obtained their results using a slightly modified version of the M-FAST. They dropped one item while continuing to rely on the recommended cut-score. In the current study, we used all the M-FAST items and the recommended cut-score.

Method

Research design

The current study used a known-groups comparison with the SIRS as a criterion for classifying forensic patients undergoing CST evaluations as either probable malingerers or nonmalingerers. Use of the known-groups comparison has a distinct advantage over simulation research in that forensic patients are facing real-world and far-reaching consequences (e.g., long prison sentences). As such, external validity is much stronger in known-groups comparisons than analogue designs (Rogers & Cruise, 1998). This study is designed to explore the efficacy of screening measures in detecting malingering in a sample of pretrial defendants undergoing competency to stand trial evaluations.

Participants

A sample of 118 forensic patients undergoing CST evaluations were recruited from a large Midwestern forensic hospital. Eighteen patients either refused to undergo psychological testing or were too impaired to complete the testing. All patients were court ordered to undergo an inpatient forensic evaluation for competency to stand trial.

The final sample consisted of 100 males ranging in age from 18 to 66 ($M = 34.26$, $SD = 11.94$). Based on self-identified ethnicities, the sample was composed of 48 African Americans, 43 European Americans, 5 Hispanic Americans, 3 Native Americans, and 1 biracial individual. The majority of the sample had a primary diagnosis of a psychotic disorder (68%); however, other disorders were also present and included mood disorders (17%), substance abuse disorders (8%), personality disorders (6%), and cognitive disorder (1%). A large majority of the sample (71%) had more than one diagnosis. In total, 80 individuals were charged with felonies and 20 with misdemeanors. The majority of the sample (70%) was charged with a violent crime and the typical patient had the potential for a very lengthy sentence ($M = 444.76$ months, $SD = 708.75$). In addition, the characteristic patient was charged with multiple offenses ($M = 3.67$, $SD = 4.29$).

Measures

The current study employed multiple measures of malingering with every patient being administered the SIRS, M-FAST, SIMS, and ECST-R ATP scale. In addition to the malingering measures, all patients were administered the Weschler Abbreviated Intelligence Scale (WASI; The Psychological Corporation, 1999) and the Evaluation of Competency to Stand Trial-Revised (ECST-R; Rogers et al., 2004). These measures were part of a competency research program and are not germane to the current investigation.

SIRS

The SIRS (Rogers et al., 1992) is a 172-item structured interview specifically designed to assess for malingering and related-response styles. It relies on eight primary scales, each focused on a specific detection strategy: Rare Symptoms, Improbable and Absurd Symptoms, Symptom Combinations, Blatant Symptoms, Subtle Symptoms, Symptom Severity, Symptom Selectivity, and Reported vs. Observed Symptoms. Each scale provides four classifications: honest, indeterminate, probable faking, and definite faking. The SIRS has demonstrated excellent psychometric

properties (internal consistency and interrater reliability) and has undergone rigorous validation (see Rogers, 2001).

M-FAST

The M-FAST (Miller, 2001) is a 25-item structured interview designed to screen for malingering. Using the SIRS as a template, Miller developed seven scales, four of which use detection strategies that are similar to the SIRS. The similar scales are Reported versus Observed (RO), Extreme Symptomatology (ES), Rare Combinations (RC), and Unusual Hallucinations (UH). Its three additional scales are Unusual Symptom Course (USC), Negative Image (NI), and Suggestibility (S). As a screen, the M-FAST has demonstrated strong psychometric properties (see Jackson et al., 2005) with an alpha coefficient for the total score of .91 and a mean inter-item correlation of .30.

SIMS

The SIMS (Widows & Smith, 2004) consists of 75 true-false questions that form five scales: Psychosis (P), Neurologic Impairment (NI), Affective Disorder (AF), Amnestic Disorders (AM), Low Intelligence (LI), along with the total score. One of the primary advantages of the SIMS is that its scales are intended to assess different facets of feigned psychopathology and cognitive impairment.

ECST-R ATP

The ECST-R ATP is composed of 28 items and was administered at the end of the ECST-R. The ECST-R ATP is organized into four scales: Rational (R), Psychotic (P), Nonpsychotic (N), and Impairment (I). Three scales (ATP-R, ATP-P, and ATP-N) are scored on a three-point scale (0 = no, 1 = sometimes, and 2 = yes). In contrast, the ATP-I is scored categorically (no or yes) to assess the reported impairment on endorsed ATP items that putatively impair the defendant's capacity to participate in the subsequent trial.

Procedure

This study is a known-groups comparison that used the SIRS as the criterion measure for categorizing patients as probable malingerers and nonmalingerers[3] (Rogers et al., 2002). Using standard cut scores, the probable malingering group consisted of defendants with three or more scales in the probable feigning range or 1 or more scales in the definite feigning range. This classification method results in very few false positives. No participants in the current study had intermediate classification (one or two elevations in the probable feigning range). Therefore, the remaining defendants were designated as nonmalingerers.

All testing was conducted in a private room where staff could visually observe patients for security reasons, but their communications were private and confidential. Data collection continued for 18 months until 100 patients were administered the complete battery (i.e., SIRS, SIMS, M-FAST, and ECST-R ATP). All patients were tested by either an experienced

[3] Relying on a single test for determining malingering should be avoided in clinical practice; instead, multiple methods (e.g., testing, direct observation) should be used to diagnose malingering.

forensic psychologist or an advanced graduate student under close supervision of a forensic psychologist.

All patients tested in the current study provided written consent for psychological testing and research upon admission into the forensic hospital. In addition, all patients were provided a verbal "forensic warning" prior to beginning the psychological testing. The forensic warning included the following information: (a) the results of testing would be used in a psychological evaluation for the court and were not confidential, and (b) the patient could decline testing or stop at any time. As such, testing conditions were similar to forensic evaluations conducted across the United States. The study was approved by the Institutional Review Boards of both the forensic hospital and the University of Wisconsin Madison.

Results

Sample characteristics

Based on the SIRS classification (see Methods), patients were categorized into two groups: 79 nonmalingerers and 21 probable malingerers. Regarding ethnic differences, African Americans (14.0%) were more likely than European Americans (4.0%, $\chi^2 = 5.64$, $p = .02$) to be classified in the probable malingering group. With respect to age, the differences were nonsignificant, $F(1, 98) = 1.47$, $p = .15$ between the probable malingering ($M = 31.29$, $SD = 9.86$) and nonmalingering ($M = 35.05$, $SD = 12.36$) groups. Likewise, the groups were comparable ($F[1, 98] = 1.20$, $p = .23$) in their potential sentences expressed in months: probable malingering ($M = 588.71$, $SD = 1056.70$) and nonmalingerers ($M = 390.74$, $SD = 535.56$), with a trend toward patients classified as malingering to be facing longer sentences. Regarding primary diagnoses, no differences existed for psychotic disorders between the probable malingering (52.4%) and nonmalingering (70.1%, $\chi^2 = 2.57$, $p = .11$) groups. The probable malingering group was more likely to have patients with a primary diagnosis of a personality disorder (23.8%) when compared to the nonmalingering (2.5%, $\chi^2 = 11.54$, $p < .01$) group. Also, the probable malingering group (19.0%) was more likely to be composed of patients with a primary substance abuse diagnosis when compared to the nonmalingering (5.1%, $\chi^2 = 4.41$, $p = .03$) group. Only one individual was diagnosed with a mood disorder in the probable malingering group so statistical comparisons between groups were not made.

Reliability of malingering screening measures

We computed alpha coefficients and mean inter-item correlations for scale and total scores. As summarized in Table 1, the malingering screens vary in their scale homogeneity. For example, the M-FAST demonstrated excellent internal consistency ($\alpha = .91$) with expected inter-item correlations ($M = .29$). Likely affected by the small number of items, alphas for most individual scales tended to be low (i.e., $<.70$). These data suggest that the M-FAST total score be used rather than its individual scales.

The SIMS yielded excellent estimates for scale homogeneity with superb data on its total score (i.e., $\alpha = .96$; M inter-item $r = .22$). Its scales demonstrated good internal consistency with the exception of Affective Disorders (Af) scale. The poor inter-item correlations ($M = .09$) contributed to its poor internal consistency ($\alpha = .61$). Focusing on the ECST-R, its ATP scales yielded good estimates of scale homogeneity. Its lowest alpha is for the ATP-Realistic scale (.72), which serves primarily as filler items and is not used to screen for feigning. All the ECST-R ATP scales demonstrated good internal consistency (α range from .72 to .90) and MICs (range from .22 to .42).

Table 1 Internal consistency and mean inter-item correlations for three malingering screens

Measure/scale	No. of items	Alpha coefficient	*M* inter-item correlation
M-FAST Total[a]	25	.91	.29
RO	3	.54	.28
ES	6	.70	.28
RC	7	.58	.16
UH	5	.68	.30
SIMS Total	75	.96	.22
N	15	.88	.33
Af	15	.61	.09
P	15	.88	.32
LI	15	.76	.17
Am	15	.93	.46
ECST-R ATP[b]			
R	9	.72	.22
P	9	.86	.42
N	9	.76	.26
I	19	.87	.27
B	18	.90	.31

Note. For M-FAST scales, RO: Reported vs. Observed; ES: Extreme Symptomatology; RC: Rare combination; UH: Unusual Hallucinations; USC: Unusual Symptom Course; NI: Negative Image; and S: Suggestibility. For SIMS scales: N: Neuologic; Af: Affective Disorders; P: Psychotic Disorders; LI: Low Intelligence; and AM: Amnestic Disorder. For the ECST-R ATP scales: R: Rational; P: Psychotic; N: Nonpsychotic; I: Impaired; B: P + N scales; and BI: B + I scores.

[a]The M-FAST also contains three scales that posses only one item. As such they were not included in the current analyses.

[b]There is no total malingering score based on the ECST-R ATP total score.

Discriminant validity of malingering screens

Consistent with past research, the current findings (see Table 2) demonstrated a high level of discriminant validity with all total and individual scales evidencing significance ($ps < .001$) and large to very large effect sizes. In our meta-analytic work on MMPI-2 (Rogers et al., 2003), we used the following conservative descriptive terms based on Cohen's d: $\geq .75$ for "moderate," ≥ 1.25 for "large," and ≥ 1.75 for "very large."[4] Applying these terms to the current data, SIMS and ECST-R ATP scales are characterized by very large effect sizes. With one exception for each measure (i.e., SIMS Af = 1.51; ATP-N = 1.81), the remaining effect sizes easily exceed 2.00. The highest effect size was found for the SIMS total score ($d = 3.07$).

The M-FAST demonstrated a very large effect size for its total score ($d = 2.69$) and three scale (ES, RC, and UH). The remaining scales had large effect sizes with the exception of USC (.85). These results are exceptionally good, given the severely restricted range for several of the M-FAST scales.

Utility of recommended cut scores

Rogers (2001) described two main reasons for using utility estimates: (a) determining the meaning of a score to a particular case and (b) assessing the overall usefulness of a measure. For

[4] Cohen (1988) recommended the following Cohen's d classifications: .2 = small, .5 = moderate, .8 = large.

Table 2 Discriminant validity in total and scale scores on the malingering screens

	Nonmalingering group ($n = 79$)		Feigning group ($n = 21$)			
	M	*SD*	*M*	*SD*	*F*	Cohen's *d*
M-FAST Total	2.38	3.14	11.76	4.60	119.82	2.69
RO	.18	.42	1.10	1.04	37.06	1.53
ES	.71	.95	2.86	1.31	71.55	2.08
RC	.41	.84	2.67	1.32	84.86	2.36
UH	.50	.88	2.38	1.32	61.15	1.91
USC	.19	.86	.86	.36	49.17	0.85
NI	.25	.44	.81	.40	27.69	1.30
S	.03	.15	.48	.52	45.98	1.66
SIMS Total	14.94	9.35	45.33	11.77	156.49	3.07
N	2.51	2.61	9.43	3.59	98.15	2.43
Af	4.94	2.38	8.52	2.32	38.18	1.51
P	2.46	2.58	9.57	2.96	118.49	2.67
LI	2.57	1.79	7.52	3.19	88.16	2.30
Am	2.46	3.16	10.29	4.01	90.46	2.34
ECST-R ATP						
P	1.06	1.84	8.95	5.11	128.47	2.79
N	1.67	2.46	6.95	4.25	54.48	1.81
I	1.14	1.68	6.81	4.69	79.36	2.18
B	2.72	3.89	15.90	8.04	114.03	2.62
BI	3.87	5.38	22.71	12.54	106.69	2.54

Note. For all *F* ratios, $p < .001$. For M-FAST scales, RO: Reported vs. Observed; ES: Extreme Symptomatology; RC: Rare combination; UH: Unusual Hallucinations; USC: Unusual Symptom Course; NI: Negative Image; and S: Suggestibility. For SIMS scales: N: Neuologic; Af: Affective Disorders; P: Psychotic Disorders; LI: Low Intelligence; and AM: Amnestic Disorder. For the ECST-R ATP scales: R: Rational; P: Psychotic; N: Nonpsychotic; I: Impaired; B: P + N scales; and BI: B + I scores.

malingering screens, the overarching goal is to ensure that very few malingerers are missed, while the large majority of non-feigners are screened out. Malingering screens perform optimally when sensitivity and NPP approach 1.00. In this study, we employed the recommended cut scores for all scale and total scores to assess their accuracy in identifying possible feigners for a complete evaluation of malingering.

As demonstrated in Table 3, the M-FAST recommended cut total score (i.e., ≥ 6 for possible feigning) performed exceptionally well (sensitivity = 1.00, NPP = 1.00). No possible malingerers would have been missed. As noted in Table 3, the total M-FAST produced good specificity (.90) and moderately high positive predictive power (.72). These data also demonstrate that the M-FAST should be used for its stated purpose as a screen because if used in lieu of a comprehensive assessment, its false-positive rate (i.e., 1-Specificity) is unacceptably high (10%) especially in light of the probable malingering rate (i.e., 21%) found in the current study. The M-FAST scales also produced good utility estimates, but do not appear to offer any advantage over the total score.

The SIMS total score produced impressive utility estimates as a screen (sensitivity = 1.00, NPP = 1.00) and missed no "probable malingerers." However, the SIMS is generally less effective than the M-FAST with PPPs less than .50. This means that more genuine patients will be screened in assessing cases of possible malingering. Nonetheless, the SIMS has a potential advantage over other malingering screens in its breath of coverage that entails possible cases of feigned amnesia and cognitive abilities.

Table 3 Testing recommended cut scores for screening measures

Subscale	Cut score	Sensitivity	Specificity	PPP	NPP	Hit rate
M-FAST Total	≥ 6	1.00	.90	.72	1.00	.92
RO	≥ 1	.66	.82	.50	.90	.79
ES	≥ 2	.81	.85	.59	.94	.84
RC	≥ 2	.81	.90	.68	.95	.88
UH	≥ 2	.67	.90	.64	1.00	.85
SIMS Total	>14	1.00	.65	.43	1.00	.72
N	>2	.95	.66	.43	.98	.72
Af	>5	.90	.71	.45	.97	.75
P	>1	1.00	.51	.35	1.00	.61
LI	>2	.90	.57	.36	.96	.64
Am	>2	.90	.67	.42	.96	.72
ECST-R ATP						
P	>1	.95	.73	.49	.98	.78
N	>0	.90	.52	.33	.95	.60
I	>1	.81	.78	.50	.94	.69
B	>2	.95	.66	.74	.98	.72
BI	>5	.90	.77	.51	.97	.80

Note. For M-FAST scales, RO: Reported vs. Observed; ES: Extreme Symptomatology; RC: Rare combination; UH: Unusual Hallucinations; USC: Unusual Symptom Course; NI: Negative Image; and *S*: Suggestibility. For SIMS scales: N: Neuologic; Af: Affective Disorders; P: Psychotic Disorders; LI: Low Intelligence; and AM: Amnestic Disorder. For the ECST-R ATP scales: R: Rational; P: Psychotic; N: Nonpsychotic; I: Impaired; B: P + N scales; and BI: B + I scores. This table is contingent upon a 21% base-rate. Of course, if that rate changes, the utility estimates change.

The ATP scales screen for several domains of feigned disorders including psychotic and nonpsychotic presentations. In the current investigation, The ATP-Psychotic (ATP-P) was the most effective scale, which missed only one probable malingerer (sensitivity = .95, NPP = .98). Three other scales including the ATP-Nonpsychotic (ATP-N; sensitivity = .90, NPP = .95), ATP-Both (Combined ATP-P and ATP-N; sensitivity = .95, NPP = .98) and Both + Impairment (Combined ATP-P and ATP-Impairment; sensitivity = .90, NPP = .97) reached the .90 benchmark for malingering screens.

An additional feature of the ATP scales is the provision of ancillary data on the type of feigning by specifically evaluating whether the defendant is feigning incompetency to stand trial. This is determined by looking at scores on the ECST-R, an established measure of competency. In this case, although classified as malingering on the SIRS, two individuals scored in the "competent" range as measured by the ECST-R (Rogers et al., 2004). While simply descriptive, this finding is helpful in cautioning forensic clinicians that they should not immediately assume that the goal of malingering is always the referral issue.

Discussion

The current investigation augments prior studies of malingering screens with a known-groups comparison that emphasizes external validity. Used with quasi-random selection, this approach provides a useful estimate regarding the prevalence of *probable* malingering at 21.0%. Given that malingering should be assessed via a multi-method approach (Rogers et al., 1992), the true prevalence rate is likely somewhat lower. Nonetheless, both current and prior studies suggest

the need in high-volume settings for effective screens to reduce the number of comprehensive assessments of malingering. The current study provides optimism that malingering screens can accurately identify cases of possible malingering and by "screening in" those in need of a more comprehensive assessment of malingering.

Each of the malingering screens examined in the current study has its respective advantages. The current data suggest that the M-FAST is best conceptualized as a total scale, because of (a) the limitations in scale homogeneity for its individual scales and (b) the lower utility estimates for these scales in comparison to the total score. The M-FAST total score proved to be very effective at identifying cases of possible malingering in the current sample. Its PPP of .74 is evidence of its effectiveness with this population.

The M-FAST is heavily weighted with items that appear to have psychotic content. A useful question is whether the M-FAST would maintain its effectiveness across a spectrum of disorders. In a known-groups comparison, we have no way to ascertain which disorder, if any, probable malingerers are likely to feign. Data from Table 2 raise the interesting question about whether fabrication of psychotic symptoms was a frequent goal. For both the SIMS and ATP, psychotic scales have the highest effect sizes (SIMS-P = 2.67; ATP-P = 2.79).

One strength of the SIMS is that its individual scales appear to have homogeneity. While their usefulness in specifying the type of feigned presentation requires further study, these scales provide forensic clinicians with useful insights into what potential domains of feigning require further evaluation. The one exception is the Af scale, which has modest internal consistency in the current research; however it demonstrated stronger internal consistency ($\alpha = .86$) in the test manual (Widows & Smith, 2004). While less effective than the M-FAST (i.e., lower PPP) with feigned mental disorders, the SIMS is effective at identifying possible cases of malingering. In the current study, we did not examine its usefulness with feigned cognitive impairment. It is possible that its breadth of coverage compensates for its lower effectiveness.

The ECST-R ATP scales appear to have good scale homogeneity and excellent discriminant validity. In the current research, we found scales that included psychotic content (ATP-P and ATP-B) were the most effective. As mentioned with the M-FAST, this finding could reflect the frequency of fabricated psychotic symptoms. Therefore, it is reassuring to see that other scales (ATP-N and ATP-I) also have very large effect sizes and are generally effective at identifying possible cases of malingering. The ATP scales differ from the M-FAST and SIMS in their focus on forensic populations, specifically those standing trial. While the ATP scales might be used in other pretrial criminal contexts, they should not be used with either nonforensic or post-trial forensic cases. When used in competency cases, they provide valuable normative data and established cut scores. In contrast, the M-FAST and SIMS are not forensic measures per se, although their results are certainly relevant to pretrial criminal cases.

Forensic clinicians are not immune to pressures within their own institutions to streamline their forensic assessments and experts should resist any pressure to substitute screens for comprehensive assessments. The false-positive rates (1-Specificity) for the current study, subject to cross-validation, range from .10 to .49. Let us assume the 21% prevalence is confirmed and that the *lowest* false-positive rate (i.e., 10%) is cross-validated. With 21 malingerers and 79 genuine patients, 8 genuine patients would be misclassified. Even this best-case scenario (i.e., high prevalence and lowest false-positives) yields unacceptable misclassifications. Avoiding Type 1 errors should be a priority when assessing malingering so genuine patients are not denied needed mental health treatment.

The overriding goal of forensic clinicians is to ensure that the judge or jury receives the most accurate information prior to rendering the verdict. Consistent with that goal, an important caveat is that the presence of malingering does not rule out genuine disorders (see Lewis et al., 2002; Jackson et al., 2005). Experts should not discontinue their evaluations once malingering

is determined. By using collateral sources, they can often establish the presence of bona fide symptoms and impairment; even in individuals who have profiles consistent with malingering.

An unexpected result of the current study was the finding that a higher proportion of African Americans were classified as malingering by the SIRS. In addition, all participants in current study were males. Future research should focus on ethnic and gender differences using known-groups design to explore and expand upon the current study. Subsequent research also should evaluate the overlap of psychopathology and malingering to improve malingering assessments with mentally disordered offenders.

In closing, malingering screens are likely to serve a valuable function in forensic assessments as they evolve and are further validated during the next decade. Presently, professional responsibility is needed to ensure that malingering scales are not misused as proxies for comprehensive assessments and are only applied to settings and populations for which they are validated.

References

American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed. Text revision). Washington, DC: Author.

Bonnie, R. J., & Grisso, T. (2000). Adjudicative competency and youthful offenders. In T. Grisso & R. G. Schwartz (Eds.), Y*outh on Trial: A developmental perspective of juvenile justice* (pp. 73–103). Chicago, IL: University of Chicago Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

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*, 387–396.

Guy, L. S., & Miller, H. A. (2004). Screening for malingered psychopathology in a correctional setting: Utility of the Miller-Forensic Assessment of Symptoms Test. *Criminal Justice and Behavior*, *31*, 695–716.

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

Lally, S. J. (2003). What tests are acceptable for use in forensic evaluations? A survey of experts. *Professional Psychology: Research and Practice*, *34*, 491–498.

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.

Luis, J. (Unpublished doctoral dissertation). *Efficacy of the structured inventory of reported symptoms (SIRS): A meta-analysis*. Univ Microfilms International.

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

Rogers, R. (1997). *Clinical assessment of malingering and deception* (2nd ed.). New York: The Guilford Press.

Rogers, R. (1999). Models of feigned mental illness. *Professional Psychology*, *21*, 182–188.

Rogers, R. (2001). *Handbook of diagnostic and structured interviewing*. New York, NY: The Guilford Press.

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

Rogers, R., & Bender, S. D. (2003). Evaluation of malingering and deception. In A. Goldstein (Ed.), *Handbook of psychology: Forensic psychology* (Vol. 11, pp. 109–129): John Wiley & Sons, Inc.

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

Rogers, R., Jackson, R. L., Sewell, K. W., & Harrison, K. S. (2004). An examination of the ECST-R as a screen for feigned incompetency to stand trial. *Psychological Assessment*, *16*, 139–145.

Rogers, R., Jackson, R. L., Sewell, K. W., & Salekin, K. L. (2005). Detection strategies for malingering: A confirmatory factor analysis of the SIRS. *Criminal Justice and Behavior*, *32*, 511–525.

Rogers, R., Salekin, R. T., Sewell, K. W., Goldstein, A., & Leonard, K. (1998). A comparison of forensic and nonforensic malingerers: A prototypical analysis of explanatory models. *Law and Human Behavior*, *22*, 353–367.

Rogers, R., Sewell, K. W., Grandjean, N. R., & Vitacco, M. (2002). The detection of feigned mental disorders on specific competency measures. *Psychological Assessment*, *14*, 177–183.

Rogers, R., Sewell, K. W., Martin, M., & Vitacco, M. J. (2003). Detection of feigned mental disorders: A meta-analysis of the MMPI-2 and malingering. *Assessment*, *10*, 160–177.

Rogers, R., Tillbrook, C. E., & Sewell, K. W. (2004). *Evaluation of Competency to Stand Trial-Revised (ECST-R) and professional manual*. Odessa, FL: Psychological Assessment Resources, Inc.

Rogers, R., & Vitacco, M. J. (2002). Forensic assessment of malingering and related response styles. In B. Van Dorsten (Ed.), *Forensic Psychology: From Classroom to Courtroom* (pp. 83–104). New York, NY: Kulwer Academic Press.

Skeem, J. L., Golding, S. L., Berge, G., & Cohn, N. B. (1998). Logic and reliability of evaluations of competence to stand trial. *Law and Human Behavior*, *22*, 519–547.

Veazey, C. H., Hays, J. R., Wagner, A. L., & Miller, H. A. (2005). Validity of the miller forensic assessment of symptoms test in psychiatric inpatients. *Psychological Reports*, *96*, 771–774.

Widows, M. R., & Smith, G. P. (2004). *SIMS: Structured Inventory of Malingered Symptomatology professional manual*. Odessa, FL: Psychological Assessment Resources, Inc.