Rule-Out and Rule-In Scales for the M Test for Malingering: A Cross-Validation

Glenn Paul Smith, PhD; Randy Borum, PsyD; and John A. Schinka, PhD

Previous research found the M test to have limited utility for the screening of malingering. Subsequently, Rogers et al. attempted to improve the test's discriminative ability by developing an alternative scoring procedure—Rule-In and Rule-Out scales. These scales showed promising results as a brief screener for malingering with hit rates as high as 95 percent. The present study cross-validated their proposed decision rules, but found lower rates of classification accuracy. The most conservative decision rule (i.e., to maximize detection of malingerers) only identified 72.7 percent of the malingerers with a false positive rate of 50.8 percent.

Within the past 10 years there has been a surge of interest in the detection of malingering. Still, there are relatively few instruments that have been developed for this purpose. However, because the issue of distinguishing real psychopathology from feigned mental illness has such profound implications for clinical and forensic assessment, the development of such tests continues to be an important area for research.

Rogers et al. has developed the Structured Interview of Reported Symptoms (SIRS), which has consistently shown promising results for the detection of malingering, even when used with actual malingerers. However, there is still a need for a briefer screening instrument. One of the first efforts to address this need was the M Test developed by Beaber et al. to detect malingering of schizophrenia.

The M Test is a 33-item inventory that comprises three separate scales. The Confusion scale is composed of eight items that are used to assess comprehension and appropriate responding. These items do not relate to malingering or symptoms of mental disorder. The Schizophrenia scale contains 10 items that correspond to the DSM-III diagnostic criteria for schizophrenia, and the Malingering scale has 15 items that would be expected to be endorsed only by malingerers. These items assess bizarre and unusual symptoms such as atypical hallucinations, atypical delusions, or extreme severity of complaints.
The psychometric properties of the instrument have also been examined. Gillis et al.\(^5\) performed a principal components factor analysis with varimax rotation that resulted in a three factor solution supporting the test's scale structure. A malingering factor accounted for the largest amount of variance. These results provided basic evidence for the construct validity of the test.

Gillis et al.\(^6\) also computed KR-20 coefficients for each of the scales to examine internal consistency. The results were as follows: Confusion scale \(r = .87\); Schizophrenia scale \(r = .87\); and Malingering scale \(r = .93\). A subsequent study of internal consistency of the scales was reported by Smith and Borum,\(^6\) who found acceptable levels of homogeneity for the Schizophrenia (alpha = .79) and Malingering (alpha = .89) scales in a forensic sample. The coefficient for the Confusion scale (alpha = .63) was felt to be somewhat low.

In the original study, Beaber et al.\(^5\) developed M test cutoff scores that correctly identified 87 percent of the patients and 78.2 percent of the simulated malingers (students instructed to fake mental illness). However, subsequent research using actual malingers showed much lower accuracy rates. Identification rates reported by Gillis et al.\(^6\) for their sample of psychiatric patients, simulated malingers, and actual malingers, however, were not as promising. Using the suggested cutting scores, they correctly identified 86.1 percent of the patients and 79.8 percent of the simulated malingers, but only 40 percent of the actual malingers. A subsequent discriminant analysis identified 93.1 percent of the patients and and 28 percent of the malingers.

The study reported by Smith and Borum also yielded identification rates that were much lower than those reported by Beaber et al. In a sample of male prisoners referred for forensic evaluation, including a group of actual malingers, the cutoff scores correctly identified 66 percent of the nonmalingers and 69.6 percent of the actual malingers. Attempts to develop a better score/scale combination to increase effectiveness were also unsuccessful. These two studies pointed out the limitations in analogue (simulation) research and suggested that the M Test in its present form had limited utility for the detection of malingering.

Subsequently, Rogers et al.\(^7\) attempted to make some improvements in the M Test to enhance its effectiveness as a screener for malingering. They developed two new scales from the existing items, called the Rule-In scale (to identify possible malingers) and the Rule-Out scale (to eliminate bona fide patients). The Rule-In scale was constructed by determining which items had the highest positive predictive power, and the Rule-Out scale used those items with the greatest negative predictive power. Each scale is composed of the 10 items with the highest predictive power for their respective scale. Internal reliability was assessed using KR-20 coefficients; both were within an acceptable range (Rule-Out \(r = .85\) and Rule-In \(r = .87\)).

Rogers et al. then developed optimal
Rule-Out and Rule-In Scales

cutting scores and provided two sequential rule options, termed options A and B. Option A uses a Rule-Out score of $<4$ followed by a Rule-In score of $<2$. These sequential rules eliminated 83.8 percent of the actual patients and retained 81 percent of the malingerers. Option B also uses a Rule-Out score of $<4$, however, this is followed by a Rule-In cutoff score of 0. This option eliminated 70.6 percent of the patients from further consideration, but retained 95.2 percent of the malingerers. Option A was recommended for use in general clinical practice. Option B was suggested for forensic settings where the index of suspicion and base rates for malingering are higher.

The present study attempted to cross-validate the two decision rules proposed by Rogers et al. using a forensic sample.

**Method**

The sample consisted of 85 male prisoners from a city/county jail who were referred for pretrial forensic evaluation. Twenty-three of these subjects had received a DSM-III diagnosis of malingering or had strong indication of malingering in their forensic reports. These subjects were defined as malingerers. There was no indication of malingering for the remaining 62 inmates. There were no significant differences between the two groups with regard to age, race, years of education, or possession of degrees/diplomas/GED.

Prior to their participation, each subject signed a consent form and release of information. They were then adminis-

tered the M Test and a demographic questionnaire.

**Results and Discussion**

Alpha coefficients were calculated to assess the internal consistency of the Rule-Out and Rule-In scales. Both were slightly lower than those reported by Rogers et al., but clearly within an acceptable range (Rule-Out alpha = .78 and Rule-In alpha = .81).

The Rule-In scale showed a significant correlation with racial status ($r = .39, p < .001$); however, there were no other significant correlations between either of the two scales and demographic variables of age, race, or education.

Using the decision rule proposed for Option A, the scales eliminated 65.6 percent of the nonmalingering prisoners from further consideration, while retaining 68.2 percent of the malingerers. Using the decision rule for Option B, only 49.2 percent of the nonmalingering prisoners were eliminated; however, 72.7 percent of the malingerers were retained.

Thus, Option A missed 31.8 percent of the malingerers and retained 34.4 percent of the nonmalingering prisoners, for a total error rate of 34 percent. While Option B only missed 27.3 percent of the malingerers, it also retained about half (50.8 percent) of the non-malingerers resulting in a total error rate of 45 percent. For each of these decision rules, the rate of error (false positives and false negatives) exceeded the 26.5 percent base rate for malingering in the sample.

This was not the case in the sample studied by Rogers et al., in which the base rate of malingering was 23.5 per-
cent. Their application of the decision rules resulted in an total error rate of 16.9 percent for Option A and 23.5 percent for Option B. While the total rate of error for Option B was equal to the base rate for malingering, only one of the malingerers escaped detection. This is also the rule suggested for use in forensic settings, where greater emphasis is placed on identifying malingerers because the false positives would probably be ruled out through further screening.

Meehl and Rosen (8) noted that one measure of diagnostic utility for a sign or score is its ability to classify individuals better than base rates alone. For example, if two percent of the population has a fear of water, and there is a Water Fear Prediction Scale which has 90 percent accuracy, the scale would not be diagnostically useful. To classify 100 people as having water fear or no water fear, it would be better to ignore the test and state that none have water fear, as this would only result in two errors (2 percent base rate). Using predictions from the scale would result in 10 errors (90 percent accuracy). Accordingly, infrequent events (i.e., those with low base rates) are more difficult to predict.

However, when using this guideline, it is also important to consider the relative costs of false negatives (misses) versus false positives. In assessing suicide risk, for example, one may be willing to accept more false positives, because the cost of a miss is very high. This same “cost analysis” operates for detection of malingering in a forensic population.

Based on the classification rates in the present sample, the Rule-In and Rule-Out scales do not appear to function as an effective screener. While we agree that there is need for brief screening measures of malingering, the M Test in its original or modified form appears to have limited utility for this purpose. It may, however, be useful to raise the index of suspicion for malingering or to provide support for a hypothesis in the presence of other clinical and/or psychometric evidence such as interview data, psychological testing, and/or the SIRS.

References
1. Rogers R: Clinical Assessment of Malingering and Deception. New York, Guilford, 1988