Have We Forgotten the Base Rate Problem?
Methodological Issues in the Detection of Distortion

Barry Rosenfeld
Long Island University

Stephen A. Sands
New York University School of Medicine

Wilfred G. Van Gorp
Cornell University Medical College

The search for valid and reliable methods of detecting malingering and distortion has become an increasingly important task for forensic psychologists and neuropsychologists. This report highlights several important methodological issues commonly observed in research on the prediction of malingering. The choice of indices for determining optimal cutoff scores on the utility of existing measures, the impact of base rates of malingering on the accuracy of prediction models, the incremental accuracy of combining multiple measures, and the relationship of test validity to the interpretation of data are described with regard to the prediction of malingering on neuropsychological tests. These methodological concerns are discussed in reference to several recent publications assessing the utility of various methods for the detection of malingering. © 2000 National Academy of Neuropsychology. Published by Elsevier Science Ltd

Keywords: malingering, neuropsychology, WMS, memory

The search for valid and reliable methods of detecting malingering and distortion has become an increasingly important task for the forensic psychologist and neuropsychologist. Whether these methods have focused on exaggeration or fabrication of psychiatric symptoms or cognitive impairment associated with a brain injury, clinicians and research scientists have continued to develop and offer methods for assessing the signs related to these important questions. Unfortunately, in the desire to quantify the likelihood of distortion, researchers have sometimes failed to carefully examine the underlying core assumptions implicit in their analysis and quantification of the likelihood of malingering.
Specifically, issues of base rates, selection criteria, and inter-test correlations have often been overlooked in much of the emerging research on detection of malingering and distortion of test results. Although many, if not most, studies suffer from these methodological issues, we will illustrate these issues with a representative example.

A recent study by Mittenberg, Azrin, Millsaps, and Heilbronner (1993) analyzed the utility of the Wechsler Memory Scales-Revised (Wechsler, 1987) for detecting exaggeration of cognitive impairment. These authors found that a basic difference score (General Memory Index [GMI] – Attention-Concentration Index [ACI]) could be used to detect malingering of cognitive impairment based upon the assumption that overall memory ability should be equal to, or superior to attention/concentration abilities. They asserted that significantly worse attentional abilities than general memory “does not make neuropsychological sense.”¹ The authors concluded, on the basis of a discriminant function analysis, that a cutoff of GMI-ACI score difference of 35 points or more provided the optimal discrimination between subjects who intentionally fabricated neuropsychological impairment from head injured subjects with genuine impairment. Along with the results of their analyses, these authors presented a table indicating the likelihood of malingering given various raw score differences between these two index scores assuming a base rate of malingering of 50%. Using this base rate of malingering (generated by contrasting 39 genuine patients to a matched sample who were instructed to malinger cognitive impairment), the authors found impressive rates of sensitivity and specificity, along with adequate rates of positive and negative predictive accuracy. However, as the base rates differ from 50%, the predictive accuracy will necessarily differ. In many clinical settings it is unlikely to be as high as 50% for most clinical settings. Indeed, Rogers and colleagues (Rogers, Salekin, Sewell, Goldstein, & Leonard, 1998; Rogers, Sewell, & Goldstein, 1994) surveyed a large sample of forensic clinicians with regard to their estimates of the base rate of malingering in their settings and found an average rate of roughly 17%, with considerable variation across estimates.

Similar methodological challenges have confronted other investigations (e.g., Iverson, Franzen, & McCracken, 1994; Prigatano & Amin, 1993), suggesting that the “base rate problem” that has frequently been discussed in several areas of research (e.g., prediction of violence or criminal recidivism, assessment of diagnostic accuracy; see Carey & Gottesman, 1978; Mossman & Hart, 1996; Shah, 1978) has largely been overlooked in this area of research. In fact, a review of the existing literature on malingering in cognitive and neuropsychological assessments yielded only two studies that directly addressed the base rate issue (Heaton, Smith, Lehman & Vogt, 1978; Mittenberg, Rotholc, Russell, & Heilbronner, 1996), calculating different estimates of predictive accuracy based on several alternative base rates.

This report seeks to highlight the importance of base rate information in understanding the accuracy of prediction methods to detect feigning, as well as identifying other methodological issues that should be considered when addressing the literature on detection of malingering.

**OPTIMIZING THE CORRECT INDEX OF PREDICTIVE ACCURACY**

Although studies of malingering have become increasingly sophisticated over the past several decades, many researchers have failed to differentiate among the various in-

¹The authors cite a secondary source as the origin of this quote but nevertheless offer this statement as theoretical justification for their method.
dices of predictive accuracy used in validation studies. In general, most classification research has focused on four indices of predictive accuracy: sensitivity, specificity, positive predictive accuracy (PPA), and negative predictive accuracy (NPA), although relatively few studies utilize all four of these indices. Using the assessment of malingering as an example, these indices reflect the proportion of malingerers who are detected by a prediction method (sensitivity), the proportion of nonmalingerers (honest respondents) who are correctly classified by the model (specificity), the proportion of individuals predicted to be malingerers who actually are malingering (positive predictive accuracy) and the proportion of individuals predicted to be honest who are actually honest (negative predictive accuracy) (see Table 1).

The importance of each of these indices of accuracy varies depending on the type of prediction offered. In the case of malingering, one could argue (as we would and Rogers, 1997, has) that positive predictive accuracy is the most clinically relevant index to examine in assessing the efficacy of a predictive model given the implications of a determination of “malingering.” An incorrect classification of a genuinely impaired individual as a malinger may lead to the unjustified denial of medical treatment and/or financial compensation for a bona fide disorder in which the individual is legitimately unable to resume work or support him or herself. Likewise, in a criminal context, a severely impaired individual may be subject to prosecution (possibly even facing a death sentence) despite neurocognitive deficits that result in an inability to understand legal proceedings or collaborate with an attorney, resulting in an unfair and biased trial. Of course, the failure to identify malingering when malingering is, in fact, present is also problematic, and may lead to unjustified compensation from civil litigation or a failure to prosecute a competent and presumably responsible defendant. The latter risk seems the lesser of the two injustices, however, because in the many cases where actual malingering is not detected the harm or consequences are not permanent. Defendants who are found incompetent to stand trial may subsequently be detected as malingerers once hospitalized for treatment of the condition associated with their incompetence. Civil litigants who successfully feign brain damage and win large financial awards might even be subject to prosecution for fraud or civil litigation themselves if they are later found to have maledgered their cognitive disorder.

Conversely, one might argue that other indices are more appropriately optimized, depending on the arena or context in which predictions are being made. For example, in

---

2Although similar views regarding the importance of positive predictive accuracy have been offered by Rogers and colleagues, this perspective has not been uniformly accepted (e.g., Mossman & Hart, 1996).
assessing the ability of clinicians or actuarial models to predict potential dangerousness of an individual, one would likely wish to optimize test sensitivity. In the case of risk assessment, the danger and damages from failing to detect potentially violent individuals represents a far greater harm than that inflicted by the needless hospitalization and treatment of a nonviolent individual. This is not to suggest that involuntary hospitalization of nonviolent individuals is a harmless loss, but it is less likely to result in permanent damage to the affected party. While other scenarios exist in which still other indices may be the most appropriate to optimize, many researchers attempt to optimize all of these indices, essentially presuming that each is equally important. This is often a challenging, if not impossible, undertaking, and may not reflect the context in which predictions are being made. Alternatively, some researchers focus exclusively on one or two indices of predictive accuracy, essentially ignoring other, potentially relevant indices.

The failure to distinguish among different indices of predictive accuracy is discussed in the review of neuropsychological methods to assess malingering by Rogers and his colleagues (Rogers, Harrell, & Liff, 1993). These authors criticize tests that utilize a “Floor Effect” to classify malingering (identifying those individuals whose performance falls below that of genuinely impaired individuals) because of their poor sensitivity (failure to identify many malingerers). Yet such tests (e.g., the Rey 15-Item Memory Test or Portland Digit Recognition Test) often have strong positive predictive accuracy ratios in many settings, indicating that individuals classified as malingering are almost invariably genuine malingerers when used in appropriate settings and with appropriate clinical samples. Thus, while a test such as the Rey 15-Item Memory Test yields a high rate of false negatives (low negative predictive accuracy), this shortcoming can be addressed by including other, more sensitive tests of malingering in the assessment process. In fact, their high PPA ratio suggests that these tests are highly accurate in accomplishing their stated objective, detection of some types of malingering, but are only effective for a particular subset of malingerers (i.e., unsophisticated). Rather than attempting to identify tests which predict malingering in all subject populations, developing instruments that maximize PPA in particular settings, and perhaps combining such tests with other tests of malingering with different accuracy and predictive properties (see below), has significant potential for the accurate assessment of malingering.

THE IMPACT OF BASE RATES ON PREDICTIVE ACCURACY

The importance of base rates in determining the accuracy of predictive models is well-established. Beginning with Meehl and Rosen’s (1955) classic paper, “Antecedent Probability and the Efficiency of Psychometric Signs, Patterns, or Cutting Scores”, researchers have consistently acknowledged that the accuracy of prediction models varies tremendously according to the base rate of the condition being predicted in any given setting. Surprisingly, however, most research on malingering has failed to examine the influence of base rates in estimating the adequacy of proposed prediction algorithms. Mittenberg et al. (1993) obtained a sample where the base rate of malingering was deliberately set at 50% by studying 39 malingering subjects and 39 matched control subjects. The authors determined the optimal cutoff score on their malingering index and their published accuracy statistics at first appear to support their recommendation of the use of this index as a viable method for detecting malingering. In fact, they report a sensitivity index of .77 (77% of malingering subjects were accurately captured) and a specificity index of .90 (only 10% of honest subjects were incorrectly classified as malingering). Although the authors did not report positive and negative predictive accuracy statistics,
TABLE 2  
Predictive Accuracy as a Function of the Base Rate

<table>
<thead>
<tr>
<th>Base Rate of Malingering</th>
<th>Predicted</th>
<th>Actual</th>
<th>Honest</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original sample—Base rate of malingering: 50%</td>
<td></td>
<td>30</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Predicted</td>
<td>Malingering</td>
<td>Honest</td>
<td>Sensitivity</td>
<td>.77</td>
</tr>
<tr>
<td>Base rate of malingering: 30%</td>
<td></td>
<td>231</td>
<td>72</td>
<td>69</td>
</tr>
<tr>
<td>Predicted</td>
<td>Malingering</td>
<td>Honest</td>
<td>Sensitivity</td>
<td>.77</td>
</tr>
<tr>
<td>Base rate of malingering: 15%</td>
<td></td>
<td>115</td>
<td>87</td>
<td>35</td>
</tr>
<tr>
<td>Predicted</td>
<td>Malingering</td>
<td>Honest</td>
<td>Sensitivity</td>
<td>.77</td>
</tr>
<tr>
<td>Base rate of malingering: 10%</td>
<td></td>
<td>77</td>
<td>92</td>
<td>23</td>
</tr>
<tr>
<td>Predicted</td>
<td>Malingering</td>
<td>Honest</td>
<td>Sensitivity</td>
<td>.77</td>
</tr>
<tr>
<td>Base rate of malingering: 1%</td>
<td></td>
<td>8</td>
<td>102</td>
<td>2</td>
</tr>
<tr>
<td>Predicted</td>
<td>Malingering</td>
<td>Honest</td>
<td>Sensitivity</td>
<td>.77</td>
</tr>
</tbody>
</table>

Note. PPA = Positive predictive accuracy; NPA = negative predictive accuracy.

these can be calculated from the data reported,\textsuperscript{3} and are .88 and .80 respectively. As can be seen, these classification ratios indicate good predictive accuracy, however, a change in the base rate of malingering will greatly alter these statistics, even with the same rates of specificity and sensitivity (see Table 2).

\textsuperscript{3}Since the authors report samples of 39 malingering and 39 nonmalingering subjects, these sample sizes are simply multiplied by the specificity and sensitivity ratios to indicate the actual number of malingerers and honest subjects that were correctly predicted. This method was then applied to other hypothetical samples, with different proportions (\(n\)) of malingering and honest subjects (Table 2), to generate PPA and NPA estimates for different base rates of malingering.
Rogers and colleagues (1993), in their review of the literature on malingering and neuropsychological assessment, estimate the base rate of malingering in most neuropsychological assessment settings to be approximately 15%. Of course, the base rate will vary across settings, with forensic settings having higher rates of attempted malingering than nonforensic settings. With a hypothetical 15% base rate of malingering and a sample size of 1000, we would expect 150 individuals to be exaggerating or feigning impairment (malingering) and 850 who are not. Applying this base rate to the data published by Mittenberg et al. (1993), with a sensitivity of .77, we would accurately detect 115 of these 150 malingerers, with 35 individuals incorrectly classified as honest despite malingering. However, with a specificity of .90, we would find that 87 of the 850 honest subjects would be incorrectly classified as malingerers. Thus, the positive predictive accuracy of this test, defined as the likelihood of correct decisions when a prediction of malingering is offered, would be only .57 (115 correct predictions out of 202; negative predictive accuracy increases to .96). In other words, almost half (43%) of the subjects who are classified as malingering are in fact responding honestly. This results in an acceptably high rate of erroneous allegations of malingering. If the base rate of malingering in a given setting is less than 15%, the PPA drops even further (see Table 2). With a base rate of malingering at 10% (a base rate likely encountered in many clinical settings), the PPA is reduced to 46%. Further, if the base rate were only 1% (an admittedly extreme scenario for many settings), the PPA would be only 7% (93% of subjects classified as malingerers would in fact be responding honestly).

Thus, although a prediction model such as that proposed by Mittenberg et al. (1993) initially appeared to adequately discriminate malingerers and nonmalingerers, it is clear that the efficacy of this model is related to the base rate of malingering in a given setting. In fact, Mittenberg and colleagues offer a probabilistic model based on the results of their discriminant function analysis in which difference scores (GMI − ACI) are used to estimate the likelihood of malingering. Such a model is both appealing and logical, however, these data are also dependent on the base rate of malingering. An inspection of this model indicates that a raw score of 0 (no difference between GMI score and ACI) corresponds to a 50% likelihood of malingering, a reasonable scenario only if the base rate of malingering were truly 50%. As the difference score increases, the likelihood estimate increases proportionately yet may still be within the “normal” range of scores (and differences among scores) on these scales. As can be seen, the failure to acknowledge the influence of base rate information may result in unjustifiably confident assertions regarding the likelihood of malingering in a given case.

**PROBLEMS ASSOCIATED WITH THE USE OF MULTIPLE TESTS**

Concern that single tests will produce incorrect results is often countered by assertions that the use of multiple tests to assess malingering will reduce the likelihood of false-positive findings. This assertion, however, may not be accurate. The use of multiple malingering measures is appealing and often necessary (given the high rates of NPA as-

---

4 Mossman and Hart (1996) argue against dichotomous classifications of malingering versus honest, however, this classification simplifies the demonstration of base rate effects.

5 Mossman and Hart (1996), in their review of the dissemination of malingering data, recommend presenting the probability of malingering for a range of possible scores and criticize Rogers et al. (1993) for their "autocratic" choice of an optimal cutoff score to base assessments of malingering. However, the data published by Mittenberg et al. (1993), demonstrate the potential for misinterpretation and misuse of probabilistic data.
associated with some measures of malingering), but may lead unsuspecting clinicians to overestimate the actual benefits provided by the use of multiple tests. Specifically, the failure to recognize when rules of probability apply and when they do not may lead to erroneous conclusions regarding the accuracy of assessment procedures.

Consider a hypothetical situation in which a neuropsychologist, in an effort to minimize the risk of error, utilizes two measures of malingering, each of which has positive predictive accuracy rates of .80 and .75, respectively. The clinician, perhaps facing a difficult cross-examination, may be asked the likelihood of incorrectly classifying an honest individual as malingering using this pair of malingering tests. The clinician will likely reply that the likelihood is less than one in 20 or .05 (i.e., by multiplying the probabilities of incorrect classifications with each test, .20 and .25, together). This statement, however, is only true when the results of each classification are independent of one another. In other words, a condition in which having been incorrectly classified as malingering on the first test does not increase the likelihood of being incorrectly classified on the second test. Such an assumption is likely incorrect. Although we are not aware of any research addressing the inter-test correspondence (i.e., reliability) of different malingering classification methods, related research has been conducted with the Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989). Greene (1997) noted correlations between several MMPI-2 measures of exaggeration ranging from .58 to .92 in nonclinical samples and .82 to .96 for patients with mental disorders. Moreover, the same type of condition which might result in an erroneous appearance of malingering on one test is also likely to be detected as malingering by other tests. Thus, in an extreme case (where all subjects incorrectly classified by one test are incorrectly classified by the second), the overall likelihood of incorrect classification may be no less than the lower of the two rates of false positives. Given the complexities that arise when multiple tests are incorporated, we will offer a simplistic example of the integration of two tests under assumptions of complete independence and near-complete overlap.

In addition to the independence of tests, a second factor influences the accuracy of multiple test interpretation: whether the clinician requires a trait or finding to be present in both versus either test before interpreting that finding as present. Consider the following scenarios based on the hypothetical test data described above. If the tests are com-

<table>
<thead>
<tr>
<th>TABLE 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classification Accuracy Using Multiple Tests</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Completely independent tests:</th>
<th>Test 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test 2</td>
<td>Correct (.75)</td>
</tr>
<tr>
<td>Correct (.80)</td>
<td>.60</td>
</tr>
<tr>
<td>Incorrect (.20)</td>
<td>.15</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Completely correlated tests:</th>
<th>Test 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test 2</td>
<td>Correct (.75)</td>
</tr>
<tr>
<td>Correct (.80)</td>
<td>.75</td>
</tr>
<tr>
<td>Incorrect (.20)</td>
<td>.00</td>
</tr>
</tbody>
</table>
pletely independent from one another, then four outcomes are possible (see Table 3): both tests might correctly identify the individual as honest (.80 \times .75 = .60), test 1 identifies the respondent as honest but test 2 suggests malingering (.75 \times .20 = .15), test 1 suggests the respondent has malingered but test 2 does not (.25 \times .80 = .20), or both tests indicate malingering (.25 \times .20 = .05). The conservative clinician, who requires both indices to be elevated may reduce the likelihood of a false allegation of malingering to only 5% by using these two tests. However, the clinician who suggests that malingering may be present when either scale is elevated has raised the likelihood of a false allegation to 40%. Although no research exists to date with regard to actual clinical practice, there is little doubt that many clinicians interpret findings from one test even when the results of other tests of the same trait or behavior provide inconsistent results. This practice may be justified when PPA rates for each test are quite high and NPA rates are relatively low (i.e., tests are highly accurate when malingering is detected but fail to detect many actual malingerers) and the incremental decrease in PPA is minimal, but in scenarios such as that described above, interpretations offered on the basis of a single test from a larger battery will likely lead to unacceptably low levels of positive predictive accuracy.

Equally striking findings might emerge when the assumption of independence is violated (as is common with tests measuring the same trait or behavior). If the condition that results in an incorrect classification of an honest subject as malingering on one test results in the same conclusion on a second test, little improvement can be expected through the use of multiple tests. This situation is actually beneficial for the clinician who concludes a person is malingering when any one of several tests is elevated, since the likelihood of inconsistent results is relatively small (i.e., when one test score is elevated others are likely to be as well). Thus, using the example above, the actual likelihood of an error may be as low as 25% (or the higher of the two false positive rates, Table 3). But for the conservative clinician who minimizes the likelihood of false positive allegations by offering interpretations only when the results of multiple tests converge, estimates of classification accuracy are substantially inflated. If, for example, the subjects incorrectly classified by hypothetical tests 1 and 2 are essentially the same subjects, then the likelihood of both scales being elevated may be as great as 20% (as low as the lowest rate of false positives). Consequently, adding multiple tests sometimes produces relatively little practical benefit in reducing the risk of false-positive assertions of malingering.

Given that the overlap among tests of malingering is unknown, the only accurate answer to questions regarding the possibility of incorrect classification would be to offer a range of possible likelihood estimates. For the conservative clinician, who requires that both tests to indicate malingering before rendering this opinion, the likelihood of an incorrect classification could range from 5 to 20%, depending on the degree of independence among these tests. On the other hand, the likelihood that a clinician who interprets either test as evidence of malingering has erred could range from 25 to 40%. Without research specifically designed to assess the degree of independence and/or overlap in classification accuracy across different tests, more accurate information regarding predictive accuracy cannot be offered.

THE IMPACT OF SUBJECT GROUP ON CLASSIFICATION ACCURACY

In the study by Mittenberg and colleagues (1993), the authors base their hypothesis that a discrepancy between the ACI and the GMI indicates probable malingering on the assumption that such a discrepancy “does not make neuropsychological sense” (p. 39). This may or may not be true, depending on the clinical issue under investigation and
population studied (Mittenberg et al. focused on patients with documented head injury). Consider a recent study by Johnstone, Erdal, and Stadler (1995), in which the authors studied the validity of the Wechsler Memor Scale-Revised (WMS-R) ACI as a measure of attention and concentration. The authors used the ACI-GMI discrepancy to classify subjects into two groups, those with probable attention deficit/hyperactivity disorders (ADHD), in which the ACI was at least 15 points below the GMI, and those without any evidence of ADHD, in which the ACI was at least 15 points greater than the GMI. The mean difference between ACI and GMI scores for the ADHD group was actually 25, a discrepancy that would suggest a 85% likelihood of malingering if the formula by Mittenberg et al. (1993) had been applied to this cohort. Of course, Mittenberg et al. offer their malingering formula as a method of detecting feigned head injury, however, the possibility that other forms of genuine cognitive impairment might mimic the performance of malingerers is problematic.

Schretlen, Brandt, Krafft, and Van Gorp (1991) demonstrated a similar effect of subject group in their study of the Rey 15-Item Memory Test. These authors observed that a large percentage of patients with genuine amnestic disorders or psychotic disorders obtained scores below the recommended cutoff for classifying malingering, despite apparently putting forth honest efforts. The authors concluded that tests such as the Rey 15-Item Memory Test may not be a valid tool for assessing malingering in some populations or settings, despite being a valid predictor of malingering in other situations.

**DISCUSSION**

The methodological issues highlighted above have often received too little in the literature on malingering and deception as it pertains to neuropsychological practice. However, given the important implications inherent when the label “malingering” is applied to a medical/psychiatric patient or forensic client, careful attention should be focused on proposed measures for the assessment of malingering. Thus, the need to consider issues such as base rates, choice of indices for assessing efficacy and the population to which the score is to be applied reflects an important issue for clinicians and researchers who seek to improve their methods of diagnostic classification.

**Implications for Research**

The concerns highlighted above have implications for the research questions and study methodologies used to identify malingering in neuropsychological testing. For example, research focused on estimating the base rate of malingering in various populations and settings can help refine estimates of predictive accuracy for tests of malingering. Mossman and Hart (1996) suggest that authors generate estimates of predictive accuracy based on both high and low estimates of the prevalence of malingering in order to adequately outline the range of predictive efficiency (as demonstrated by Mittenberg et al., 1996). While we would agree with this suggestion, further research is necessary to generate accurate estimates of these parameters. In addition, studies designed to assess the independence or overlap among different measures of malingering can help refine estimates of predictive accuracy when multiple assessment techniques are integrated. Only through careful attention to the methodological issues involved in prediction research, as well as a clear exposition of these issues and limitations when publishing one’s findings, can authors insure an accurate representation of their data.
Implications for Clinical Practice

A number of clinical issues also arise from this discussion that warrant recapitulation. First, given the obvious interdependence of base rates and classification accuracy statistics, clinicians should be particularly cautious when offering estimates of the likelihood of malingering. While summary tables such as that published by Mittenberg et al. (1993) provide considerable temptation for clinicians in forensic settings, without data on the base rate of malingering in particular settings such estimates are likely to be incorrect. Likewise, while incorporating multiple tests of malingering can be a useful method of overcoming the limitations of any individual tests, the method by which such tests are combined and the manner in which findings are conveyed is crucial to a fair and accurate representation of the data. Without empirical evidence of the extent to which different malingering test findings correlate with one another, clinicians should likely offer a range of possible probability estimates and/or provide a range of possible error rates which accurately reflects both the ambiguity of the data as well as the potential for misclassification. With these caveats, clinicians can more accurately convey the true extent of their knowledge and acknowledge the limitations in their assessment methods. Given the importance of fairness in legal proceedings as well as clinical work, these goals appear not only desirable, but virtually mandatory.

REFERENCES


