Interpreting Clinical Evidence of Malingering: A Bayesian Perspective

Douglas Mossman, MD

Customary ways of reporting on or testifying about malingering have shortcomings. Stating an opinion "with reasonable medical certainty" tells fact-finders little about how much confidence the opinion deserves; stating that an individual's behavior is similar to that of known malingerers does not convey the information that fact-finders really need to know, which is the likelihood that the evaluee in question is a malingerer, given the evaluator's findings. Mossman and Hart (Mossman D, Hart KJ: Presenting evidence of malingering to courts: insights from decision theory. Behav Sci Law 14:271–91, 1996) recommend that mental health professionals address this problem by using Bayes' theorem to interpret test data from evaluations. However, these authors do not discuss the use of evidence obtained during interviews and from other clinical contexts, nor do they describe a method for quantifying imprecision in Bayesian probabilities. This article provides examples of how forensic evaluators might use a Bayesian perspective to interpret clinical indicia of malingering observed during evaluations of adjudicatory competence. The article discusses sources of imprecision in Bayesian posterior probabilities, describes a method for characterizing that imprecision using confidence intervals, and then presents several sample calculations that illustrate how interview findings change the likelihood of malingering. The article also discusses the implications of the Bayesian approach for forensic evaluations and for future research on malingered incompetence.

J Am Acad Psychiatry Law 28:293–302, 2000

Criminal defendants often pretend to have psychiatric problems. It is estimated, for example, that more than 30,000 evaluations of adjudicatory competence are performed annually in the United States,[1] and studies and surveys suggest that roughly one-sixth of competence-to-stand-trial (CST) evaluees feign incapacitating mental disorders.[2–4] Taken together, these reports imply that each year, U.S. forensic evaluators see more than 5,000 defendants who are feigning incompetence to stand trial.

Two types of evidence help forensic evaluators detect malingering. Evaluators sometimes use test data from psychological assessment tools that either include validity scales (e.g., the MMPI-2)[5] or have been designed specifically to distinguish malingerers from honest responders (e.g., the Structured Interview of Reported Symptoms (SIRS)[6] and the VIP (Validity Indicator Profile)[7]). Evaluators also use less formal (but nonetheless valuable) pieces of evidence derived from their observations of evaluees, interviews, documents, and other available history. Several authors suggest that these clinical indicia of malingering (e.g., absurd or atypical symptoms, calling attention to one's illness, or inconsistencies between reported problems and observed behavior) are useful clues for identifying evaluees who feign or exaggerate mental illness.[8–10]

When mental health professionals interpret and present their findings on malingering to courts, they typically do so in one of two ways. When the basis of the opinion is primarily clinical data, professionals often state simply that they believe an evaluee was or was not malingering and describe the findings supporting their opinion. When the opinion is based on test data, professionals may report that evaluee's responses are "consistent with persons who honestly report symptoms" or are "characteristic of persons who are malingering" (Ref. 6, p. 24).

These interpretive approaches can generate problems and misunderstandings, however. Either kind

Dr. Mossman is Professor and Director, Division of Forensic Psychiatry, Wright State University School of Medicine, Dayton, OH, and Adjunct Professor, University of Dayton School of Law. Address correspondence to: Douglas Mossman, MD, WSU Dept. of Psychiatry, P.O. Box 927, Dayton, OH 45401-0927. E-mail: dmossman@pol.net

of statement misleadingly suggests that clinical indicia or test results give yes-or-no answers about the presence of malingering, whereas the true effect of diagnostic information is to refine our beliefs about how likely it is that a condition is present.[11] A professional's statement of belief about malingering tells the fact-finder little about how sure the professional is or about how much confidence an opinion deserves, even when the professional also gives good reasons for the opinion. Statements that an individual's behavior is similar to that of known malingerers may have better scientific grounding than categorically expressed opinions, but such statements may be objectionable in some jurisdictions (e.g., Washington state).[12]

Physicians should easily identify a final shortcoming with these interpretive approaches. When doctors perform a diagnostic test, they are not concerned with how often people who are already known to have a particular disorder have a positive test result. Instead, what doctors want to know is the inverse of this, "the probability that any particular test result, positive or negative, is a true result" (Ref. 13, p. 1019). Similarly, telling judges, lawyers, or jurors whether an evaluee has behaved like a malingerer does not convey the information that attorneys and fact-finders really need to know, which is the likelihood that the evaluee is malingering, given the evaluator's findings.

Mossman and Hart[14] have recommended that mental health professionals address these problems by using Bayes' theorem[15] when interpreting data about malingering and presenting their opinions on malingering in court. For many years, scholars have recommended that clinicians and investigators use Bayes' theorem to express their level of confidence concerning a broad range of hypotheses.[16] Bayesian reasoning is a staple of medical decision-making literature,[17] and the *Annals of Internal Medicine*[18] recently encouraged investigators to interpret their results using Bayesian methods. Bayesian reasoning also has been recommended as a vehicle for interpreting psychological diagnoses[19] and the results of criminal investigations,[20] as well as for presenting evidence in court.[21, 22]

Using published results from three studies of malingering scales and tests, Mossman and Hart[14] explain how Bayes' theorem could be applied; they also suggest that reanalyses of already existing investigations might let professionals give courts well-founded probabilistic estimates of malingering in a variety of forensic contexts. Because Mossman and Hart focus on the interpretation of test data, they leave open the question of whether evaluators could use Bayesian principles to interpret clinical data relevant to malingering. This question is important because not all evaluees undergo systematic testing for malingering and because formal testing for malingering is often performed because clinical findings suggest that such testing is needed. As Rogers and Salekin[4] point out, Mossman and Hart also do not address the problem of describing and characterizing how imprecision in data and estimated frequencies affects the calculation of Bayesian probabilities.

This article shows how forensic evaluators might use a Bayesian perspective to interpret observations and interview data obtained during evaluations of adjudicatory competence. The article first describes Bayes' theorem as it applies to this evaluation context. Next, the article discusses sources of imprecision in Bayesian posterior probabilities and describes a method for characterizing that imprecision using confidence intervals, thereby addressing a major concern raised by Rogers and Salekin.[4] The article then presents several calculations that illustrate how clinical findings might affect an evaluator's belief that an evaluee is malingering. Finally, the article examines the implications of the Bayesian approach for forensic evaluations and future research on malingered incompetence.

Bayes' Theorem

Suppose that a clinical sign, T, is associated with malingering and can take on n different values. We can use the symbol T_i to denote a particular value of T, where $i = 1,2,\ldots,n$. Let $M+$ denote the presence of malingering and $M-$, the absence of malingering. The base rate, or pretest probability of malingering, is represented by the symbol $P(M+)$. For those clinical signs that are binary (i.e., either present or absent), we can avoid subscripts and use the simple notation $T+$ (positive test, sign present) and $T-$ (negative test, sign absent). If T is a binary sign, we can also use the familiar terms sensitivity and specificity to describe its "diagnostic performance" in detecting malingering. Sensitivity and specificity are thus conditional probabilities. Sensitivity is the probability that sign T will be present, given that the evaluee is malingering, and is represented symbolically as $P(T+|M+)$. Specificity is the probability that

T will be absent, given that an evaluee is not malingering, and is represented symbolically as $P(T-|M-)$.[11]

Suppose, next, that an evaluee displays clinical sign T, that is, he is $T+$. How does the presence of T change what we believe about the probability that he is malingering? Bayes' theorem tells us that the new probability is:

$$P(M+|T+) = \frac{P(T+|M+)P(M+)}{P(T+|M+)P(M+)+[1-P(T-|M-)\,[1-P(M+)]} \quad \text{(Eq. 1)}$$

Equation 1 states that $P(M+|T+)$, the conditional probability of the evaluee's malingering given the presence of T, is a function of the pretest probability of malingering $P(M+)$ and the sensitivity $P(T+|M+)$ and specificity $P(T-|M-)$ of T as a sign of malingering.

For nonbinary signs (those that can have more than two values), we write Bayes' theorem as follows:

$$P(M+|T_i) = \frac{P(T_i|M+)P(M+)}{P(T_i|M+)P(M+)+P(T_i|M-)[1-P(M+)]} \quad \text{(Eq. 2)}$$

Equation 2 states that $P(M+|T_i)$, the conditional probability of the evaluee's malingering when T has the value i, is a function of three things: the pretest probability of malingering $P(M+)$, the probability that a malingerer's value of T will equal i, and the probability that a nonmalingerer's value of T will equal i.

A numerical example can help illustrate how clinical information (here, the evaluee's being $T+$) alters an evaluator's estimate of the probability of malingering. Cornell and Hawk[23] compared the frequency of several purported signs of malingering in 39 pretrial evaluees who were diagnosed by experienced forensic examiners to be malingering psychotic symptoms with the frequency of these signs in 25 genuinely psychotic patients.* They found that 24 of 39 malingerers displayed "exaggerated behavior" (i.e. dramatic, unusual actions such as barking like a dog), but only 6 of 25 genuinely psychotic patients did so. (This difference is highly significant: $\chi^2 = 8.6$, $df = 1$, $p = .003$.) The sensitivity, $P(T+|M+)$, of exaggerated behavior (EB), or $P(EB+|M+)$, is 24/39 = .62 and the specificity, $P(T-|M-) = P(EB-|M-)$, is 19/25 = .76.

If a particular evaluee displays EB (i.e., if he is $T+$ or $EB+$), what is the probability that he is malingering? Assume for the moment that we know that 22 percent of CST evaluees who exhibit putative signs and symptoms of mental disability are, in fact, only feigning mental problems, so that $P(M+) = .22$; also assume that the evaluation context is similar to that used by Cornell and Hawk,[23] and that their findings about the accuracy of this clinical sign are applicable. From Equation 1, we find that a $T+$ evaluee's probability of faking is:

$$P(M+|T+) = \frac{.62 \times .22}{.62 \times .22 + (1 - .76) \times (1 - .22)} = .42 \quad \text{(Eq. 3)}$$

Although Cornell and Hawk report that rates of exaggerated behavior in malingerers and genuine patients differ significantly, finding exaggerated behavior only justifies a modest increase in the estimated probability of malingering.

Confidence Intervals and Posterior Probabilities

Many readers will have noticed that the values for sensitivity and specificity used in Equation 3 were derived from a small, although presumably representative, sample of evaluees. Were Cornell and Hawk[23] to evaluate another 39 malingerers and 25 psychotic patients, we would expect results similar to their published findings but not exactly the same. This is because empirical data, even when gathered systematically from representative samples of the relevant populations, reflect (1) the true values one would obtain from testing everyone on the face of the earth, and (2) random sampling errors.

Inspecting Equation 1, we see that uncertainty in the estimate of $P(M+)$ as well as sampling error in

* In their study, Cornell and Hawk[23] classified clinical signs as being either present or absent, and much of this article therefore uses Equation 1 to examine the diagnostic properties of these signs. However, exaggerated behavior and many other indicia of malingering could be conceptualized as present to varying degrees in malingering and honest evaluees. If clinicians used, for example, a five-point rating scale to categorize evaluees' behavior, diagnostic properties would be described using Equation 2, with $i = 1, 2, 3, 4$, or 5, and five pairs of conditional probabilities.

$P(T+|M+)$ or $P(T-|M-)$ may contribute to imprecision in $P(M+\ T+)$. We would like to have a way of quantifying this imprecision (i.e., a way to construct confidence intervals that would specify a range within which the result probably falls). The following discussion explains how to characterize the imprecision in the estimates of $P(M+)$, $P(T+|M+)$, and $P(T-|M-)$ using beta-distributions and how to use these distributions to construct a reliable confidence interval for $P(M+\ T+)$.

Confidence Interval for P(M+), the Prior Probability or Base Rate

A Bayesian framework forces clinicians to formulate their clinical experience, often educated "hunches," in mathematical terms. This requirement may also encourage clinicians to find support for those hunches by gathering data, especially if those hunches must be defended in court. This subsection uses concrete examples of plausible data to illustrate how one might generate a mathematical characterization of one's preevaluation knowledge about malingering by pretrial evaluees.

Such knowledge might come from two sources, which we can informally designate as "background information" and "empirical data." An example of the former is results summarized by Rogers and Salekin,[4] who report that the expert consensus on the prevalence of malingering is .1744 ± .1444. An example of the latter is the findings reported by Gothard and colleagues,[2] who, in examining 55 CST evaluees, found that 25 were competent, 23 were incompetent, and seven were malingering. Testing for malingered incompetence is concerned with distinguishing those who are feigning illness from those who are truly incompetent. Therefore, the relevant portion of Gothard and colleagues' sample does not include the 25 defendants who were deemed competent but only the 30 evaluees who appeared incompetent, who might have been faking incompetence, and among whom seven were found actually to be malingerers.

Suppose Gothard and colleagues[2] (or other investigators who thought these data were applicable to their work setting) wished to use this background information and empirical data to generate a mathematical characterization—what statisticians call a "probability distribution function"—of their knowledge about the probability of malingering in CST evaluees who appeared incompetent. Iverson[24] and Kleiter[25] suggest that such knowledge can be described using the beta-distribution, which utilizes this general formula:

$$f(\pi) = C \cdot \pi^{s-1}(1 - \pi)^{t-1} \qquad \text{(Eq. 4)}$$

In Equation 4, π is the unknown proportion of malingerers, s and t determine the shape of the probability distribution, and C is a "normalizing" constant that is calculated in such a way that the area under the distribution between $\pi = 0$ and $\pi = 1$ equals unity. A short-hand notation for this distribution is "$\beta(s,t)$."

The distribution described by Equation 4 has a mean μ and standard deviation σ. These are related to the shape parameters s and t in the following way:[24]

$$s = \mu\left[\frac{\mu(1-\mu)}{\sigma^2} - 1\right]$$

$$t = (1-\mu)\left[\frac{\mu(1-\mu)}{\sigma^2} - 1\right] \qquad \text{(Eq. 5)}$$

In a discussion that goes beyond the scope of this article, Iverson[24] explains how one combines the background prior distribution and the empirical data using Bayes' theorem. The result is this posterior distribution for the proportion π of malingerers given the empirical data q:

$$f(\pi)f(q|\pi) = f(\pi|q) = C \cdot \pi^{x+a-1}(1-\pi)^{n-x+b-1} \qquad \text{(Eq. 6)}$$

Notice that the right side of Equation 6 has the same form as Equation 4.

To assign values to the parameters a and b, we return to the results described by Rogers and Salekin,[4] which constituted our background information. If the expert consensus on the prevalence of malingering is .1744 ± .1444, we can let $\mu = .1744$ and $\sigma = .1444$. These values are substituted into Equation 5 to generate an initial beta-distribution with parameters a and b that describes the background knowledge. Doing the arithmetic yields results in which a is approximately 1 and b is approximately 5.

The parameters n and x reflect the empirical data;

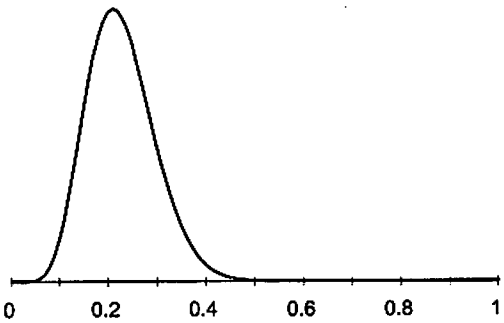

Figure 1. A probability distribution function for the preevaluation likelihood of malingering, based on data of Gothard and colleagues.[2]

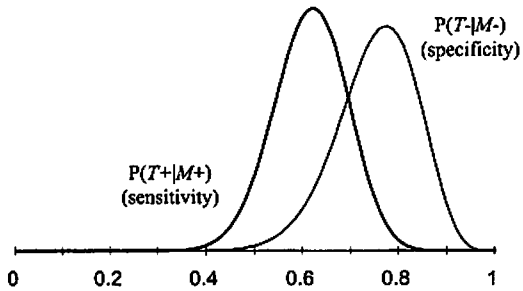

Figure 2. Probability distribution functions for the sensitivity and specificity of exaggerated behavior as an indicator of malingering, based on data of Cornell and Hawk.[23]

x is the number of malingerers (in this case, 7) and n is the total number in the sample (i.e., 30). Plugging these results into Equation 6 gives $\beta(x + a, n - x + b) = \beta(8,28)$ as the probability distribution function for the preevaluation probability of malingering. The resulting distribution is shown in Fig. 1; its mean is .222, and 95 percent of this distribution lies between .104 and .369. We thus have an example of an explicit Bayesian characterization of the base rate of malingering.

Confidence Intervals for Sensitivity and Specificity

CST evaluations may yield all sorts of clinical data that evaluators can use to distinguish malingerers from actually incompetent evaluees. For clinical findings that are either present or absent, the diagnostic accuracy of these data will be characterized by sensitivity = $P(T+|M+)$ and specificity = $P(T-|M-)$. Newcombe[26] and Kleiter[25] suggest that the same beta-distribution discussed in the previous subsection permits an efficient way of establishing confidence intervals for fractions of this sort. To return to our earlier numerical example, we can represent the sensitivity of exaggerated behavior, $P(EB+|M+)$, as $\beta(25.5,14.5)$; the specificity of exaggerated behavior, $P(EB-|M-)$, can be represented by $\beta(19.5,6.5)$.† The contours of these distribution functions can be discerned readily in Fig. 2. The 95 percent confidence intervals for $P(EB+|M+)$ and $P(EB-|M-)$ are .485 to 0.777 and .571 to 0.893, respectively.

Confidence Intervals for Post-Test Likelihoods

The considerations presented above will let us illustrate how to construct confidence intervals for the posterior probability of malingering in $EB+$ evaluees, which Equation 2 estimated as .42. Kleiter[25] has described a complicated formula for such confidence intervals using Polya-Eggenberger distributions. An equally acceptable (and much simpler) method‡ is to generate a secondary distribution for $P(M+|EB+)$ using the results of random (Monte Carlo) samplings from the beta-distributions for $P(M+)$, sensitivity, and specificity (see for example Ref. 28). In other words, one uses a computer to: (1) draw, at random, a value from the above-described β-distribution for the CST evaluees' prior probability of malingering; (2) draw, at random, one value each from the β-distributions for $P(EB+|M+)$ and $P(EB-|M-)$; (3) combine these three values in Equation 1 to obtain a value of $P(M+|EB+)$; (4) repeat this process N times, where N is a large number; (5) use the distribution of these N results to make inferences about confidence intervals for $P(M+|EB+)$.

For this article, I performed this procedure with $n = 2{,}000$. The central 1,900 of the 2,000 values fell between .184 and .688, so we can use this as the 95 percent confidence interval for $P(M+|EB+)$ in CST evaluees.

† The inclusion of an additional ½ in the parameters of these distributions is derived from "objective Bayes" theory, based on the use of objective, or noninformative, priors. The common noninformative prior for a success probability p is the $\beta(½,½)$ distribution, which, when supplemented with data involving s successes and f failures, yields the $\beta(s+½,f+½)$ posterior. $\beta(½,½)$ is the so-called "Jeffreys prior," which is defined as the square root of the Fisher information and which, in one dimension, is the best prior in terms of yielding confidence sets with good coverage.[27]

‡ In several simulation experiments, I have found that the resulting confidence intervals for $P(M+|T+)$ have coverage properties that are very close to the nominal value. That is, for example, the 95 percent confidence interval includes the "true" value approximately 95 percent of the time, is too high 2.5 percent of the time, and is too low 2.5 percent of the time (data on file with the author).

Table 1 Clinical Evidence of Malingering[a]

Clinical Indicator	Sensitivity	Specificity	Probability of Malingering
Affect flat or inappropriate	0.026 (0.003–0.114)	0.480 (0.295–0.669)	0.014 (0.001–0.074)
Loose, circumstantial, or pressured speech	0.051 (0.011–0.154)	0.200 (0.081–0.384)	0.018 (0.001–0.048)
Using neologisms	0.026 (0.003–0.114)	0.640 (0.445–0.805)	0.020 (0.002–0.106)
Incoherent speech	0.051 (0.011–0.154)	0.640 (0.445–0.805)	0.039 (0.007–0.160)
Concrete thinking	0.051 (0.011–0.154)	0.680 (0.485–0.836)	0.044 (0.007–0.180)
Poor hygiene	0.051 (0.011–0.154)	0.720 (0.527–0.865)	0.050 (0.008–0.177)
Inattentive, distractible	0.077 (0.022–0.191)	0.640 (0.445–0.805)	0.058 (0.013–0.185)
Grandiose delusions	0.103 (0.036–0.226)	0.600 (0.406–0.773)	0.068 (0.016–0.193)
Ideas of reference	0.103 (0.036–0.226)	0.640 (0.445–0.805)	0.075 (0.018–0.223)
Persecutory delusions	0.256 (0.140–0.407)	0.520 (0.331–0.705)	0.132 (0.048–0.290)
Absurd replies	0.282 (0.160–0.435)	0.880 (0.713–0.965)	0.402 (0.145–0.713)
Exaggerated behavior	0.615 (0.459–0.755)	0.760 (0.571–0.893)	0.423 (0.184–0.688)
Memory problems	0.256 (0.140–0.407)	0.920 (0.767–0.983)	0.478 (0.162–0.854)
Visual hallucinations	0.462 (0.313–0.616)	0.960 (0.828–0.996)	0.767 (0.367–0.968)
Endorsing bogus symptoms	0.205 (0.102–0.350)	1.000 (0.905–1.000)	1.000 (0.315–1.000)
Suicidal ideation	0.359 (0.223–0.515)	1.000 (0.905–1.000)	1.000 (0.450–1.000)
Symptoms fit various diagnoses	0.385 (0.245–0.541)	1.000 (0.905–1.000)	1.000 (0.472–1.000)
MMPI-2-*F*–*K* score > 29	0.840 (0.721–0.921)	0.960 (0.878–0.992)	0.857 (0.576–0.971)
SIRS (exactly three elevated scales)[b]	0.228 (0.175–0.289)	0.995 (0.976–0.999)	0.927 (0.683–0.991)

[a] Sensitivity = $P(T+|M+)$, Specificity = $P(T-|M-)$, and probability of malingering = $P(M+|T+)$ for several clinical indicia of, and two tests for, malingering. Numbers in parentheses are 95% confidence intervals. See text for explanations of prior probability and calculations.
[b] Values in this row equal $P(T_{SIRS-3}|M+)$, $1 - P(T_{SIRS-3}|M-)$, and $P(M+|T_{SIRS-3})$; see footnote (**) to text.

Evaluating and Comparing Clinical Signs

Exaggerated behavior is just one of many clinical indicators of malingering, and one might wonder how these indicators compare with each other. It turns out that the methods described in the previous section have potential usefulness not just for constructing confidence intervals related to clinical indicia, but they also give us a tool for comparing the performances of various clinical signs with each other and with more formal test methods such as the MMPI-2 and the SIRS.

To illustrate this concept, I used the above described Monte Carlo method to generate confidence intervals for the post-test probability of malingering associated with several of the clinical signs and symptoms studied by Cornell and Hawk(1989).[§] Data for the MMPI-2 *F-K* scale come from Graham and colleagues (Table 2 in Ref. 29)[¶]; data for having exactly three elevated SIRS scales come from Rogers and colleagues (Table 16 in Ref. 6).[**] The prior probability distribution used here is the $\beta(8,28)$ distribution developed in the previous section for Gothard and colleagues' data.[2]

In a frequently cited discussion of malingered psychosis, Resnick[8] writes:

> Malingerers have less success imitating the form than the content of schizophrenia. . . . With respect to the form, derailment, neologism, and incoherence are rarely simulated. . . . Malingerers are unlikely to show negative symptoms or the subtle signs of residual schizophrenia, such as impaired relatedness, blunted affect, digressive speech, or peculiar thinking. . . . Malingerers' symptoms may fit no known diagnostic entity but represent symptoms from various diagnoses [p. 60, citations omitted].

The results given in Table 1 concerning affect disturbance, disordered speech and thought, concrete thinking, grandiosity, and impaired concentration (that is, signs and symptoms that represent the "form" of psychosis) do what Resnick says they do. All reduce the likelihood of malingering and should

§ Cornell and Hawk[23] report on 24 clinical signs and symptoms. For this article, I examined data on signs or symptoms for which the *p* values fell below .05 (without a Bonferroni correction), plus memory problems, absurd symptoms, and persecutory delusions.

¶ In the data published by Graham and colleagues, *F-K* scores for the 50 honest responders and 50 subjects instructed to "fake bad" fell across a broad range of values. For this example, I have dichotomized the scores as being above, less than, or equal to 29. Scores for 42 of the "fake bad" subjects were >29, but only two of the honest responders scored that high. Note that Graham and colleagues obtained these results from 100 college students, whose test-taking behavior may have been quite different from the performance of criminal defendants undergoing forensic evaluations.

** Table 16 and the accompanying text in Rogers and colleagues' manual[6] make it possible to calculate how many feigners and honest responders had 0–2 elevated subscales, 3 elevated subscales, and 4 or more elevated subscales. One of 196 honest responders had exactly 3 elevated subscales, so $P(T_{SIRS-3}|M-) = 1/196 = .0051$; 47 of 206 malingerers had exactly 3 elevated subscales, so $P(T_{SIRS-3}|M+) = 47/206 = .228$. In Table 1 in the present work, Equation 2 was used to calculate the values for the SIRS, and Equation 1 was used for the other indicia.

therefore increase an examiner's confidence that an evaluee's reported psychosis is genuine. By contrast, suspicion of feigned psychosis is raised by a "symptom" picture that includes complaints associated with many diagnoses. Reports of suicidal ideation, agreeing to bogus symptoms suggested by the examiner, and visual hallucinations also increase the likelihood of malingering.

It is reassuring, but hardly significant, to find that Table 1 is in accord with Resnick's descriptions. What Table 1 adds to Resnick's discussion is the possibility of quantifying how much various clinical findings should influence an examiner's opinion about malingering. Suppose a forensic examiner found himself in a situation where the preevaluation probabilities assumed for this article's calculations and Cornell and Hawk's[23] finding were totally applicable. Table 1 says that certain clinical findings against malingering should let him feel fairly confident that psychosis is not being feigned. By contrast, those same pretest probabilities, coupled with sensitivities and specificities associated with clinical indicia of feigning, allow the examiner to be less confident when he encounters signs indicative of malingering. This is true even though the indicia that support a diagnosis of malingering are just as sound as the indicia that count against malingering.

Compare, for example, the implications of loose, circumstantial, or pressured speech with the implications of reported symptoms fitting multiple diagnoses. Finding the former implies that there is a 97.5 percent chance that the probability of malingering is less than .048. But finding the latter only tells the examiner that there is a 97.5 percent chance that the probability of malingering is more than .472. When an evaluee endorses symptoms that fit multiple diagnoses, this changes the posterior probability of malingering (calculated from Equation 1) from .222 to 1, but the confidence interval for this probability has a long left tail.

The other clinical findings that increase the probability of malingering—reports of suicidal ideation, bogus symptoms, and visual hallucinations—also produce posterior probability confidence intervals with long left tails. So do the two types of test data used to complete Table 1. Although the confidence intervals for $F\text{-}K > 29$ and for three elevated SIRS scales are narrower than those for the clinical findings, this does not necessarily mean that these formal tests are "better" at detecting malingering. As the first two columns of numbers in Table 1 show, the sensitivities and specificities for $F\text{-}K > 29$, SIRS = 3, and the best clinical indicia are similar. However, the confidence intervals for the tests' sensitivities and specificities are narrower than the intervals for clinical indicia because the tests' accuracies have been evaluated in larger numbers of subjects (100 subjects for the MMPI-2; 402 for the SIRS).

Of additional interest is what Table 1 tells us about how exaggerated behavior, memory problems, and absurd replies affect the probability of malingering. Cornell and Hawk[23] report that the rates of exaggerated behavior in malingerers and genuinely psychotic evaluees differed significantly, but the rates of memory problems and absurd replies did not. Despite this report, these three clinical findings have a similar impact on the estimated posterior probability of malingering and on that estimated probability's confidence intervals. What is suggested is that the meaning and usefulness of a clinical finding is not a simple function of statistical significance.

Discussion

The numerical results presented in this article are offered merely to illustrate how posterior probabilities and confidence intervals might be generated by applying a Bayesian approach to malingering data. Although these calculations use plausible assumptions and data obtained from well-reasoned, peer-reviewed articles, the specific findings described in the figures and in Table 1 should be viewed with great skepticism. The data presented by Cornell and Hawk[23] reflect only the behavior of individuals who attempted to feign psychotic symptoms; "faking dumb," which is often encountered in pretrial evaluations, was not discussed. As is often the case in malingering research, Cornell and Hawk's "gold standard" for determining whether an evaluee actually was malingering was expert clinical judgment using all available data; if this judgment was in error, it might translate into error in accuracy indices.††

†† The absence of a "gold standard" —and the potential errors this can cause —is a vexing problem in malingering research that attempts to evaluate behavior of individuals in real settings who are faced with real consequences for their behavior.[30] In the past two decades, several investigators have developed statistical techniques that can quantify diagnostic accuracy in the absence of a perfect, independent means of establishing the truth about an individual's status; such techniques may be applicable to research on the accuracy of malingering assessments. Ref. 31 provides an introduction to these statistical issues that is geared to a psychiatric audience.

Also, the calculations for clinical indicia reflect the behavior of small numbers of criminal defendants whose adjudicatory competence was evaluated at two sites and may not be applicable to other sites or to other types of forensic evaluations. The data used for calculations concerning the *F-K* index (and, to some degree, the SIRS) come from subjects who were not facing criminal charges and may not have accurately simulated the behavior of malingering and honestly responding pretrial evaluees.

Nonetheless, the preceding discussion, table, and figures illustrate the potential feasibility of Bayesian interpretations based on reexamined existing data or data from new studies. This is important, because such interpretations capture the role of clinical data as information that alters one's beliefs about the likelihood of a condition. It is intuitively obvious that, when an evaluator accumulates information about an evaluee, that information (among other things) should influence the evaluator's belief about the evaluee's honesty. By explaining quantitatively how information affects beliefs, Bayes' theorem can help evaluators figure out how important, or unimportant, clinical data are. The resulting probabilistic interpretation of clinical findings lets evaluators plan additional evaluation strategies (e.g., to request or do specific testing) and communicate the import of their findings to courts.‡‡

Adopting the Bayesian approach would require some distinct changes in how most forensic evaluators approach their work and interpret findings. Bayes' theorem depicts the logical necessity of interpreting information in light of one's prior knowledge, that is, of the prior probabilities of the phenomena under examination. Therefore, examiners who want to apply Bayes' theorem to clinical findings on malingering must make explicit estimates of their preevaluation beliefs about an evaluee probability of feigning illness. Ideally, such estimates should be context-specific, so that, for example, a preevaluation probability of malingered adjudicatory incompetence would reflect the evaluator's general knowledge about the phenomenon (e.g., information from publications) and applicable local data. Evaluators thus may need to review relevant recent experience to optimize their preevaluation estimates of malingering.

When such numerical information is not available from publications or other formal sources, forensic clinicians can still employ the Bayesian approach using probability estimates with broad confidence intervals to reflect those estimates' imprecision. Several years ago, Raiffa[40] suggested that decision-makers could use systematic self-interrogation to give their personal knowledge about pretest likelihoods a mathematical characterization. Quantifying hunches is often a requirement of the Bayesian framework. In a forensic context, Bayes' theorem makes clinicians force themselves to specify how their "experience, training," and "specialized knowledge" (Fed. Evid. R. 702) inform their judgments.

Some readers may believe that the quality and limited generalizability of data used in this article's calculations make those calculations pointless. It is important to recognize, however, that this view really is a criticism of the data used—and of using clinical indicia to aid judgments about malingering—rather than a criticism of Bayesian principles. Readers who agree that the calculations in this article provide insights concerning the usefulness of clinical indicia must also acknowledge that the Bayesian approach accentuates the need to investigate more thoroughly the mathematical features of clinical malingering data. Kucharski and colleagues[41] point out that although several authors feel that "clinical presentation variables" can be used to detect malingering, "there are few empirical data supporting the validity or accuracy of clinical presentation or history variables in differentiating defendants with *bona fide* psychiatric disorder from those feigning mental illness" (p. 580). Additional studies such as the one performed by Cornell and Hawk[23] would tell evaluators more about the frequency of various clinical indicia in malingering and honest evaluees.

For example, given the frequency with which CST evaluations are performed, it might be desirable and relatively easy to conduct multi-site studies to examine several clinical indicia in large numbers of com-

‡‡ Not all courts would view such interpretations favorably. For example, the Minnesota Supreme Court has worried that frequency statistics might be taken by jurors to imply a "quantification of the likelihood that the defendant... is guilty" (Ref. 32, p. 548); the Connecticut Supreme Court forbade use of Bayes's Theorem in a criminal case on grounds that it requires the jury to adopt a prior probability of guilt.[33] Opposition to probabilistic evidence has not arisen in the majority of cases that have raised this issue.[34–38] Many of these cases have dealt with probabilistic interpretations of DNA evidence for establishing paternity; such interpretations have been supported by the American Bar Association and the American Medical Association.[39] A full discussion of this issue would take me far beyond the scope of this article, which has suggested that Bayesian interpretations of malingering can optimize mental health professionals' understanding of their findings and their communication to fact-finders. For good or ill, courts will remain free to reject what mental health professionals have to offer, and some courts may choose to do so.

petence evaluees. Evaluating large numbers of subjects would let forensic evaluators use sensitivities and specificities with narrower confidence intervals, which would generate more precise estimates of posterior probabilities; conducting studies at several sites might help evaluators learn how well findings generalize from one evaluation setting to the next. Properly designed studies might also yield multiple-variable discriminant functions that would tell evaluators how to combine data concerning several indicators and that would have better classificatory accuracy than individual variables.[41]

Individuals vary greatly in their styles and degrees of malingering,[30] yet the probabilities generated by Equation 1 concern the presence or absence of a phenomenon, rather than degrees or styles. However, determining that an evaluee has a certain probability of malingering need not prevent an evaluator from commenting on the evaluee's degree or style of malingering. Thinking in Bayesian terms can help evaluators describe to fact-finders their confidence in rendering certain conclusions. It may then be valuable for evaluators to describe additional features of their findings that characterize in detail how an evaluee has responded to the evaluation.

Even if Bayesian conceptualizations are the most appropriate way of presenting findings, judges and jurors may not understand or make optimal use of probabilistic evidence. Some commentators (e.g., Tribe[42]) and courts have voiced concern that frequency statistics will have an overwhelming "impact on the trier of fact" (Ref. 43, p. 482).[43] However, a recent study by Smith and colleagues[44] suggests, among other things, that jurors may under-utilize Bayesian probabilities, a finding that accords with previous research.[45]

The issue of how well persons use base rates and probabilistic information is highly complicated and a source of contention among scholars.[46, 47] This issue is distinct, however, from the central point of this article, which is that clinical findings relevant to malingering alter likelihoods in a way that is fittingly described by Bayes' theorem. Mental health professionals should recognize this principle when they think about their findings and explain them to legal decision-makers. Bayesian reasoning can help mental health professionals avoid over- or under-interpreting the significance of their findings, both to themselves and, to the extent that courts will permit, to fact-finders.

Acknowledgment

The author thanks Professor James O. Berger of Duke University for suggestions concerning interval construction for posterior probabilities.

References

1. Nicholson RA, Kugler KE: Competent and incompetent criminal defendants: a quantitative review of comparative research. Psychol Bull 109:355–70, 1991
2. Gothard S, Rogers R, Sewell KW: Feigning incompetency to stand trial: an investigation of the Georgia Court Competency Test. Law Hum Behav 19:363–73, 1995
3. Rogers R, Sewell KW, Goldstein A: Explanatory models of malingering: a prototypical analysis. Law Hum Behav 18:543–552, 1994
4. Rogers R, Salekin RT: Beguiled by Bayes: a reanalysis of Mossman and Hart's estimates of malingering. Behav Sci Law 16:147–53, 1998
5. Bagby RM, Rogers R, Buis T: Detecting malingered and defensive responding on the MMPI-2 in a forensic inpatient sample. J Pers Assess 62:191–203, 1994
6. Rogers R, Bagby RM, Dickens SE: Structured Interview of Reported Symptoms (SIRS) Professional Manual. Odessa, FL: Psychological Assessment Resources, 1992
7. Frederick RI: VIP: Validity Indicator Profile Manual. Minneapolis: National Computer Systems, 1997
8. Resnick PJ: Malingered psychosis, in Clinical Assessment of Malingering and Deception (ed 2). Edited by Rogers R. New York: Guilford Press, 1997, pp 47–67
9. Cunnien AJ: Psychiatric and medical syndromes associated with deception, in Clinical Assessment of Malingering and Deception (ed 2). Edited by Rogers R. New York: Guilford Press, 1997, pp 23–46
10. Pankrantz L, Binder LM: Malingering on intellectual and neuropsychological measures, in Clinical Assessment of Malingering and Deception (ed 2). Edited by Rogers R. New York: Guilford Press, 1997, pp 223–36
11. Mossman D, Somoza E: Neuropsychiatric decision making: the role of prevalence in diagnostic testing. J Neuropsychiatry Clin Neurosci 3:84–8, 1991
12. State v. Petrich, 683 P.2d 173 (Wash. 1984)
13. Davidoff F: Standing statistics right side up. Ann Intern Med 130:1019–21, 1999
14. Mossman D, Hart KJ: Presenting evidence of malingering to courts: insights from decision theory. Behav Sci Law 14:271–91, 1996
15. Bayes T: An essay towards solving a problem in the doctrine of chances. Philos Trans R Soc Lond 53:370–5, 1763
16. Fischoff B, Beyth-Marom R: Hypothesis evaluation from a Bayesian perspective. Psychol Rev 90:239–60, 1986
17. Sox HC, Blatt MA, Higgins MC, Marton KI: Medical Decision Making. Boston: Butterworths, 1988
18. Information for authors. Ann Intern Med 127:1–15, 1997
19. Meehl PE, Rosen A: Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. Psychol Bull 52: 194–216, 1955
20. Sullivan RC, Delaney HR: Criminal investigations: A decision-making process. J Police Sci Admin 10:335–43, 1982
21. Fienberg SE, Schervish MJ: The relevance of Bayesian inference for the presentation of evidence and for legal decision making. Boston U L Rev 66:771–98, 1986
22. Matthew R: Improving the odds on justice? New Sci 142:12–13, 1994
23. Cornell DG, Hawk GL: Clinical presentation of malingerers di-

agnosed by experienced forensic psychologists. Law Hum Behav 13:375–83, 1989
24. Iverson GR: Bayesian Statistical Inference. Newbury Park, CA: Sage Publications, 1984
25. Kleiter GD: Propagating imprecise probabilities in Bayesian networks. Artif Intel 88:143–61, 1996
26. Newcombe RG: Two-sided confidence intervals for the single proportion: comparison of seven methods. Stat Med 17:857–72, 1998
27. Jeffreys H: Theory of Probability (ed 3). Oxford: Clarendon Press, 1961
28. Parmigiani G, Berry DA, Aguilar O: Determining carrier probabilities for breast cancer—susceptibility genes BRCA1 and BRCA2. Am J Hum Genet 62:145–58, 1998
29. Graham JR, Watts D, Timbrook RE: Detecting fake-good and fake-bad MMPI-2 profiles. J Pers Assess 57:264–77, 1991
30. Rogers R: Introduction, in Clinical Assessment of Malingering and Deception (ed 2). Edited by Rogers R. New York: Guilford Press, 1997, pp 1–19
31. Faraone SV, Tsuang MT: Measuring diagnostic accuracy in the absence of a "gold standard." Am J Psychiatry 151:650–7, 1994
32. State v. Kim, 398 N.W.2d 544 (Minn. 1987)
33. State v. Skipper, 637 A.2d 1101 (Conn. 1994)
34. Kaye DH: The admissibility of "probability evidence" in criminal trials: Part II. Jurimetrics J 27:160–172, 1987
35. United States v. Porter, 618 A.2d 629 (D.C. Ct. App. 1992)
36. Graham v. State, 308 S.E.2d 413 (Ga. Ct. App. 1983)
37. Davis v. State, 476 N.E.2d 127 (Ind. Ct. App. 1985)
38. Kammer v. Young, 535 A.2d 936 (Md. Ct. App. 1987)
39. Abbott JP, Sell KW, Krause HD, *et al*: Joint AMA-ABA Guidelines: present status of serologic testing in problems of disputed parenting. Fam Law Q 10:247–88, 1976
40. Raiffa H: Decision Analysis: Introductory Lectures on Choices Under Certainty. New York: Random House, 1968
41. Kucharski LT, Ryan W, Vogt J, Goodloe E: Clinical symptom presentation in suspected malingerers: an empirical investigation. J Am Acad Psychiatry Law 26:579–85, 1998
42. Tribe L: Trial by mathematics: precision and ritual in the legal process. Harv L Rev 84:1329–93, 1971
43. State v. Boyd, 331 N.W.2d 480 (Minn. 1983)
44. Smith BC, Penrod SD, Otto AL, Park RC: Jurors' use of probabilistic evidence. Law Hum Behav 20:49–82, 1998
45. Kaye DH, Koehler JJ: Can jurors understand probabilistic evidence? J R Stat Soc Ser A 154:75–81, 1991
46. Koehler JJ: The base rate fallacy reconsidered: descriptive, normative, and methodological challenges. Behav Brain Sci 19:1–53, 1996
47. Commentary on Jonathan J. Koehler (1996). Behav Brain Sci 20:774–83, 1997