February, 1996

Weighing Evidence in Sexual Abuse Evaluations: An Introduction to Bayes's Theorem

James M. Wood, *University of Texas at El Paso*

Weighing Evidence in Sexual Abuse Evaluations: An Introduction to Bayes's Theorem

James M. Wood
University of Texas at El Paso

This article introduces readers to Bayes's theorem, a simple mathematical formula that can illuminate general issues and assist evaluators in the field of child sexual abuse. The theorem is applied to a case study of sexual abuse allegations that arose during a custody dispute.

The evaluation of child sexual abuse requires the assessment of specific facts in the light of general scientific and professional knowledge. The evidence of this particular case is placed within the framework of what is known in general. The task of applying general knowledge to a specific case can be a difficult one. For example, an evaluator may know that a particular child has exhibited sexual behaviors and that research shows such behaviors to be more common among abused than nonabused children. How is the evaluator to apply scientific knowledge to the present case? Is the evidence sufficient to warrant further investigation? To "validate" abuse? To turn the case over to the criminal justice system?

Such decisions are commonly made on the basis of professional judgment. However, professional judgment and experience can provide a poor basis for decision making (see review by Dawes, Faust, & Meehl, 1989). Human beings, including highly educated professionals, are prone to "slips of the mind" when they must evaluate situations that involve substantial uncertainty (Tversky & Kahneman, 1974; Wood & Wright, 1995), as many sexual abuse cases do.

The uncertainty of individual evaluators is reflected in broader dilemmas that confront the field of sexual abuse evaluation. How useful are behavioral indicators (e.g., nightmares, bed wetting, or sexual behaviors) in determinations regarding abuse? What credibility should be given to allegations that arise in the context of custody disputes? What are the advantages and drawbacks of different techniques for interviewing children?

The present article provides an introduction to Bayes's theorem, a 200-year-old mathematical formula that can assist evaluators and clarify important issues in the field of child sexual abuse. For instructional purposes, we begin by describing an actual custody case that involved sexual abuse allegations. At the end of this article, we will return to the case and discuss it in the light of Bayes's theorem.

CASE STUDY

Patricia B. contacted Child Protective Services to allege that her ex-husband, Alfonso B., was sexually abusing their 5-year-old son, Billy, during parental visitations. (Names are changed to protect confidentiality.)

Based on Patricia's allegations, a judge ruled that Alfonso's visits with his son should be supervised. However, Alfonso was allowed to choose his own supervisors. Immediately following one of these visits, Billy was taken to the hospital by his mother. A physical examination followed by laboratory tests revealed the presence of sperm in the boy's rectum.

Subsequently, Billy made several statements to police. He reported that on the day of the visitation, his father had sent the supervisors away and then digitally penetrated him. However, Billy also told the police that his mother had put "ointment" on his "butt" before taking him to the hospital for the physical examination.

A court-appointed psychologist, Dr. Alcoa, subsequently concluded that Patricia had put sperm on her son's bottom, apparently to substantiate allegations of sexual abuse against her ex-husband. Billy was taken from his mother and placed into foster care pending a court decision regarding custody.

CHILD MALTREATMENT, Vol. 1, No. 1, February 1996 25-36

$$\frac{\text{Percentage of abused children who imitate intercourse}}{\text{Percentage of nonabused children who imitate intercourse}} = {}^{14\%}\!/_{1\%} = {}^{14}\!/_{1} \text{ or } 14{:}1. \tag{A}$$

$$\frac{\text{Percentage of abused children who are overly aggressive or overly passive}}{\text{Percentage of nonabused children who are overly aggressive or overly passive}} = {}^{35\%}\!/_{10\%} = {}^{3.5}\!/_{1} \text{ or } 3.5{:}1. \tag{B}$$

AN INTRODUCTION TO BAYES'S THEOREM

The Strength of Evidence and the Likelihood Ratio

Sexual abuse evaluators routinely assess the strength of evidence. For example, an evaluator may offer an opinion that bed wetting constitutes "weak" evidence of sexual abuse, or that gonorrhea in a child constitutes "extremely strong" evidence. Words such as *weak, moderate,* or *strong* are commonly used to describe the strength of evidence. The same purpose may be accomplished by using a number called the likelihood ratio (LR).

> *It is not enough to know that evidence is weak or strong. Exactly the same evidence may lead to quite different conclusions, depending on the rate of abuse in the group being evaluated.*

The LR expresses the relative probability of coming across a particular piece of evidence in one group rather than in another. As an example, we may consider a case in which a child has been observed to imitate sexual intercourse. According to reports from parents (Friedrich et al., 1992), this behavior occurs in about 14% of abused children and 1% of nonabused children. The LR for this evidence may be calculated as shown in equation A above.

The LR of 14:1 indicates that imitation of intercourse is about 14 times more common among abused than nonabused children.

As a second example, consider a child who is overly aggressive or overly passive. According to parent reports (Friedrich et al., 1992), about 35% of abused children, but only 10% of nonabused children, are overly aggressive or overly passive. The LR for the evidence may be calculated as before in equation B above.

As these examples show, LRs are usually written so that one side of the ratio is 1 and the other side is greater than 1. When written in this way, two LRs may be easily compared. A glance shows that the LR for imitation of intercourse (14:1) is about four times larger than the LR for overly aggressive, overly passive behavior (3.5:1). The comparison indicates that imitation of intercourse is substantially stronger evidence of sexual abuse than is overly aggressive, overly passive behavior.

The evidentiary "strength" represented by an LR can be expressed in common English terms. In Table 1, adapted from Goodman and Royall (1988), various LRs are given along with their English "translations." As can be seen, the LR of 14:1 for imitation of intercourse translates into moderate-to-strong evidence of abuse. By contrast, the LR of 3.5:1 for overly aggressive, overly passive behavior translates into weak evidence.

Reaching Conclusions: A First Try Using the Likelihood Ratio

As discussed in the previous section, imitation of sexual intercourse by a child is about 14 times more common among abused than nonabused children. The behavior has an LR of 14:1 and constitutes moderate-to-strong evidence of abuse. Now readers are asked a tricky question. Suppose that a particular child is observed to imitate sexual intercourse. Speaking in round numbers, what is the probability that the child has been sexually abused?

After taking a moment, perhaps readers have calculated that the child is 14 times more likely to have been abused than not abused, and that the probability of abuse is therefore 14 out of 15, or about 93%. Unfortunately this answer, which is commonly given, is incorrect. We confess, with apologies, that the question was even trickier than it appeared. In fact, it could not be answered because insufficient information was provided.

In order to answer a question of this type, two things must be known beforehand: (a) the strength of the evidence that the child has been sexually abused, and (b) the rate of sexual abuse among children in the group being evaluated. Readers were provided with information about the strength of the evidence, but not about the rate of abuse. Without information about the rate, no conclusion could be drawn from the evidence.

To see how the rate of abuse can influence the evaluation of evidence, let us consider two hypotheti-

TABLE 1: Likelihood Ratios and Suggested Interpretations

Likelihood Ratio	*English Translation*
1:1	No evidence
3:1	Weak evidence
5:1	Weak-to-moderate evidence
7:1	Moderate evidence
14:1	Moderate-to-strong evidence
20:1	Strong evidence
55:1	Very strong evidence

NOTE: Based on Goodman & Royall, 1988.

cal school districts, each with 10,000 children. In the first, which we shall call the Dismal School District, about 20% of the children have been sexually abused. In the second, the Protective School District, the prevalence of sexual abuse has been reduced to about 5%. In both schools, the behavior of imitating sexual intercourse is 14 times more common among abused than nonabused children.

Now suppose that two children are seen for evaluation because they have been observed to imitate sexual intercourse. One child is from the Dismal School District, the other from the Protective School District. What is the probability that these children have been sexually abused?

Again, we hope readers have taken a moment to answer the question. This time there really is a correct answer. In fact, as a sort of compensation for the tricky question earlier, there are *two* correct answers. The probability of abuse for the child from the Dismal School District is 78%. The probability of abuse for the child from the Protective School District is 42%.

Some readers may have difficulty believing that the child from the Dismal School District is almost twice as likely to have been abused as the child from the Protective School District. After all, in both cases the evidence is the same: the child has been observed to imitate sexual intercourse. Therefore, common sense seems to argue that both children are equally likely to have been abused.

However, as can be seen in Table 2, this is one of those cases where untutored common sense leads to the wrong conclusion. In the Dismal School District, about 360 children imitate sexual intercourse. Of these 360, 280 (78%) have been abused and the rest have not. By contrast, in the Protective School District far fewer children, 165 in all, imitate intercourse. And of these 165 children, 70 (42%) have been abused and the rest have not.

A close look at the figures for the Protective School District provides insight into the situation. The number of abused children (and therefore the number of abused children who imitate intercourse) is *lower* in the Protective School District than in the Dismal School District. At the same time, the number of nonabused children (and therefore the number of nonabused children who imitate intercourse) is *higher*. The "mix" of abused versus nonabused children is different in the two schools *and* in the subgroups of children who imitate sexual intercourse.

As this example shows, it is not enough to know that evidence is weak or strong. Exactly the same evidence may lead to quite different conclusions, depending on the rate of abuse in the group being evaluated. Therefore, a thoughtful evaluator will take account of both the evidence and the rate of sexual abuse. The task of combining these two kinds of information is made easier by the mathematical formula known as Bayes's theorem. But before discussing the theorem, we must first introduce some useful concepts.

Base Rates, Prior Odds, and Posterior Odds

The terms *base rate* and *prevalence* are closely related and may be considered equivalent for the purposes of the present discussion. The base rate of something is simply the relative frequency with which that thing occurs in a particular group. For instance, research indicates that about 20% of females in the United States have been seriously sexually abused at least once (see review by Salter, 1988). The base rate (or the prevalence) of sexual abuse for this group is therefore about .20. By extension, the base rate for nonabuse is about .80.

Base rates may vary depending on the particular group being considered. For example, according to some estimates, 70% of sexual abuse allegations made to Child Protective Services are reliable (Jones & McGraw, 1987). Therefore, the base rate of abuse among such cases is .70. By extension, the base rate of "not abused" is around .30.

For purposes of decision making, base rates are often converted into *prior odds*. Prior odds are the ratio between two different base rates for the same group. For example, if the base rate of abuse among cases involving allegations to Child Protective Services is around .70, then the base rate of "no abuse" in the same group is about .30. We can therefore calculate the prior odds of abuse versus no abuse in such cases as follows:

$$\frac{\textit{Base rate of abuse}}{\textit{Base rate of no abuse}} = {}^{.70}\!/_{.30} = {}^{7}\!/_{3} \text{ or } 7{:}3. \qquad \text{(C)}$$

The prior odds of abuse versus no abuse are about 7:3. This means that if we were to select such cases at random, there would be seven reliable reports for every three that are unreliable or false.

Prior odds are different from probabilities but closely related to them. If the prior odds of abuse are

TABLE 2: The Probability That a Child Who Imitates Sexual Intercourse Has Been Sexually Abused: Example Using Two Hypothethical School Districts

Assumptions:

1. Prevalence of abuse in Dismal School District is 20%.
 Prevalence of abuse in Protective School District is 5%.
2. About 14% of abused children imitate sexual intercourse; 1% of nonabused children imitate sexual intercourse.

Dismal School District		*Protective School District*	
2,000 abused children	8,000 nonabused children	500 abused children	9,500 nonabused children
280 abused children imitate intercourse	80 nonabused children imitate intercourse	70 abused children imitate intercourse	95 nonabused children imitate intercourse
360 children in all imitate intercourse		165 children in all imitate intercourse	
78% (280/360) of all children who imitate intercourse have been abused		42% (70/165) of all children who imitate intercourse have been abused	

9:1, then the "prior probability" of abuse is .90. This probability is calculated simply by adding the two sides of the LR (9 + 1 = 10), and then dividing the left side of the LR by the resulting sum (9 ÷ 10 = .90).

By themselves prior odds can be very informative. In a sense, they summarize the way things usually are. In our daily lives, we use them all the time to assess events that we encounter. For example, if a little boy tells us that he saw a dog in the front yard, we are inclined to believe him without difficulty. But if he tells us that he saw an elephant, we are unlikely to accept his statement without question. The prior odds of an elephant are so low that we are likely to be dubious.

In sexual abuse cases, prior odds can be similarly informative. As an example, consider a mother who has made allegations of sexual abuse to a physician, who in turn has contacted Child Protective Services. As we have already noted, the prior odds of a reliable versus an unreliable allegation are probably about 7:3. These odds are known, or at least could be, before a single shred of evidence is gathered. Even before investigation, therefore, it is known that the allegation is about twice as likely to be reliable as unreliable.

But although prior odds are informative, the ultimate concern in sexual abuse cases is with *posterior odds*—the odds of abuse *after* evidence has been collected and evaluated. The difference between prior odds and posterior odds can be easily illustrated. Suppose a worker at Child Protective Services is about to investigate a sexual abuse allegation. The worker believes that about 70% of such allegations are probably true (Jones & McGraw, 1987) and that the prior odds of a true versus an unreliable or false allegation are therefore 7:3.

Suppose that the worker then interviews the child witness, who gives a logical, coherent, and detailed description of sexual abuse. After the interview, the worker concludes that the odds of a true versus a false allegation are 99:1. These odds, which are arrived at after the evidence has been evaluated, are called the posterior odds of abuse.

The distinction between prior odds and posterior odds hinges on whether the evidence has been evaluated yet. The prior odds are the odds of abuse before the evidence has been evaluated. The posterior odds represent the odds of abuse after the evidence has been taken into account.

A *posterior probability* can be calculated from the posterior odds in the same way that the prior probability was earlier calculated from the prior odds. In the example just given, the posterior probability of abuse is 99% (99 ÷ 100). In sexual abuse evaluations, the determination of posterior odds is of the utmost importance. That is, after reviewing the evidence in a case, the evaluator is expected to decide whether abuse is likely to have occurred. The evaluator may estimate the posterior odds of abuse using clinical judgment. An alternative approach is to calculate the odds using the mathematical formula known as Bayes's theorem.

Reaching Conclusions: A Second Try Using Bayes's Theorem

As discussed earlier, in order to draw conclusions from evidence, a sexual abuse evaluator must know two things: the strength of the evidence and the rate of abuse in the group to which the child belongs. The strength of evidence is represented numerically by the likelihood ratio (LR). The rate of sexual abuse is represented by the prior odds of abuse versus no abuse.

More than 200 years ago, Thomas Bayes, an English clergyman and mathematician, proposed a theorem that allows conclusions to be drawn from the LR and prior odds (Bayes, 1763; see also Fisher, 1959). This theorem may be expressed as follows:

$$\text{Prior Odds} \times \text{Likelihood Ratio} = \text{Posterior Odds}$$

The formula for Bayes's theorem shows that the posterior odds of abuse can be calculated in a straightforward manner by multiplying together the prior odds and the LR for the evidence. For example, let us consider a child from the Dismal School District who has been observed imitating sexual intercourse. The posterior probability that this child has been abused may be calculated by applying Bayes's theorem in three steps.

Step 1, the prior odds of abuse are calculated. As stated above, 20% of the children in the Dismal School District have been molested. The base rate of abuse in the district is therefore .20, the base rate of no abuse is .80, and the prior odds of abuse are .20:.80 or 1:4.

Step 2, the LR of the evidence is calculated. As already discussed, imitation of intercourse is about 14 times more common among abused than nonabused children. The LR is therefore 14:1.

Step 3, the posterior odds are calculated by multiplying together the prior odds and the likelihood ratio.

$$\text{Prior odds} \times \text{Likelihood ratio} = \text{Posterior odds}$$

$$\tfrac{1}{4} \times \tfrac{14}{1} = \tfrac{14}{4} \text{ (78\% probability of abuse)}. \qquad \text{(D)}$$

As can be seen, the posterior odds of abuse are 14/4 or 14:4. The posterior probability of 78% indicates that about three fourths of the children from the Dismal School District who are observed imitating sexual intercourse have in fact been molested. Note that 78% is the figure given earlier in Table 2.

Bayes's theorem can also be used to calculate the probability of abuse for a child from the Protective School District who has been observed to imitate intercourse. The same three steps are followed as before. Step 1, the base rate of abuse for that district is .05, the base rate of no abuse is .95, and the prior odds of abuse versus no abuse are therefore .05:.95, or 1:19. Step 2, the LR for imitation of intercourse, as before, is 14:1. Step 3, the prior odds and LR are combined using Bayes's theorem to calculate the posterior odds of abuse:

$$\tfrac{1}{19} \times \tfrac{14}{1} = \tfrac{14}{19} \text{ (42\% probability of abuse)}. \qquad \text{(E)}$$

As can be seen, the posterior odds of abuse are 14/19 or 14:19, and the posterior probability of abuse is therefore 42%, which is the same number given in Table 2.

A posterior probability, like an LR, can be translated into common English. Table 3 provides "translations" of posterior odds and posterior probabilities into ordinary language. These translations correspond with common English usage. However, they should be regarded as guidelines rather than rules.

TABLE 3: Posterior Odds as Related to Probabilities and Common English Terms

Posterior Odds of Abuse vs. No Abuse	*Probability That Abuse Has Occurred*	*Conclusion (in common English)*
1:19 or less	5% or less	Very unlikely
1:19 to 1:4	5% to 20%	Unlikely
1:4 to 2:3	20% to 40%	Somewhat unlikely
2:3 to 3:2	40% to 60%	Undetermined
3:2 to 4:1	60% to 80%	Somewhat more likely than not
4:1 to 19:1	80% to 95%	Likely
19:1 or more	95% or greater	Very likely

According to Table 3, for example, the 78% posterior probability of abuse for the child from the Dismal School District can be translated as "somewhat more likely than not" that abuse has occurred. However, 78% is so close to 80% that an evaluator might reasonably conclude that abuse is "likely." Either translation of the probability would be appropriate. Similarly, as shown in Table 3, the 42% posterior probability of abuse for the child from the Protective School District can be translated to mean that the case is of "undetermined" validity. The finding of undetermined indicates that the odds of abuse are so close to 50-50 that the evaluator really cannot say with certainty whether or not abuse has occurred. It should be noted that an undetermined finding is *not* the same as a finding of "unlikely." In fact, about 50% of the children who fall into this category have in fact been abused.

Evidence That Tends to Decrease the Posterior Odds of Abuse

So far, the discussion has focused on evidence that *increases* the belief that abuse has occurred. For example, the presence of gonorrhea in a child, or a report that a child has imitated sexual intercourse, increases the belief that abuse has occurred and the posterior

$$\frac{\text{Sexual behavior if child has been abused}}{\text{Sexual behavior if child has \underline{not} been abused}} = \frac{35\%}{5\%} = \frac{7}{1} \text{ or } 7:1 \qquad \text{(G)}$$

$$\frac{\text{Sexual behavior ABSENT if child has been abused}}{\text{Sexual behavior ABSENT if child has \underline{not} been abused}} = \frac{65\%}{95\%} = \frac{1}{1.46} \text{ or } 1:1.46 \qquad \text{(H)}$$

probability of abuse. However, some types of evidence have the opposite effect. For example, evidence that the accused perpetrator was in another city at the time of the alleged offense greatly decreases the posterior probability of abuse.

Bayes's theorem works equally well with either type of evidence. The only difference is in the form of the LR. For example, consider a case in which sexual abuse has been alleged in the context of a custody dispute. As explained later in this article, the LR of "custody dispute" can be estimated as about 1:2. Notice that in this LR, unlike those that have appeared earlier, the *1* appears on the left side of the ratio instead of the right. Thus, when this LR is multiplied times the prior odds, the posterior odds will be smaller, not larger, than the prior odds.

If this particular case of alleged sexual abuse comes from the Dismal School District, where the prior odds of abuse are 1:4 (prior probability = 20%), then the posterior odds of abuse can be estimated as follows:

$$\tfrac{1}{4} \times \tfrac{1}{2} = \tfrac{1}{8} \text{ (11\% probability of abuse).} \qquad \text{(F)}$$

As can be seen, the posterior probability of abuse (11%) is close to the prior probability (20%). The difference is small because evidence with an LR of 1:2 is weak, having precisely the same strength as an LR of 2:1 but in the opposite direction. Similarly, evidence with an LR of 1:7 is moderate, as is evidence with an LR of 7:1. An LR and its mirror image have equal but opposite strengths. Thus, Table 1 can be consulted to determine the strength of any LR.

The Absence of Evidence

In some sexual abuse cases, the evaluator must consider the absence of evidence rather than its presence. For example, a physician may have examined the child and reported no physical evidence consistent with abuse. How is the evaluator to weigh this absence of evidence? Some individuals, particularly defense lawyers, have suggested that the presence and absence of evidence should be given equal but opposite weights. For instance, if the presence of physical evidence indicates that abuse has occurred, some have argued, then the absence of physical evidence must indicate that abuse has *not* occurred.

Such reasoning is definitely incorrect. There are two ways to demonstrate the error. The first is to use a simple example. Imagine that a detective investigating a sexual abuse case happens to open a drawer and find an erotic note from the alleged perpetrator to the victim. Most people would agree that the presence of the note would constitute strong evidence against the alleged perpetrator. But suppose that the detective had opened the drawer and found it empty. Would the absence of a note constitute strong evidence that the alleged perpetrator was innocent? Clearly the answer is no. Such notes are seldom found in genuine abuse cases. There is no symmetry in this respect. The presence of a particular piece of evidence may be very weighty indeed, whereas the absence of that same evidence may have almost no significance one way or the other.

The same point may be illustrated using Bayes's theorem. Research indicates that about 35% of sexually abused preschool children show some form of inappropriate sexual behavior (Kendall-Tackett, Williams, & Finkelhor, 1993). Comparable figures for unabused children are unavailable, but 5% is probably a reasonable estimate. The LR for the presence of sexual behavior in a child can be calculated as shown in equation G above.

The LR for the absence of sexual behavior in a child can also be calculated. However, some other probabilities must be computed first. Using the figures already cited, we can determine that sexual behavior is absent in 65% (100% − 35%) of sexually abused preschool children, and in 95% (100% − 5%) of preschool children who have *not* been abused. The LR can therefore be calculated as shown in equation H above.

As can be seen, the two LRs are not equal and opposite at all. The presence of sexual behavior constitutes moderate evidence (LR = 7:1) that the child has been abused, whereas the absence of sexual behavior constitutes very weak evidence (LR = 1:1.46) that the child has *not* been abused.

APPLICATION OF BAYES'S THEOREM TO GENERAL ISSUES IN THE FIELD OF SEXUAL ABUSE

The preceding discussion has introduced the basic terminology and logic of Bayes's theorem. However, the theoretical nature of the discussion should not obscure the practical nature of the theorem. In the following section, Bayes's theorem will be used to clarify important general issues in the field of sexual abuse evaluation.

Personal Experience and the Interpretation of Evidence

Imagine two child protection workers, Worker A and Worker B. Worker A investigates all reports from the Protective School District that involve sexual behaviors in children. The prior odds of abuse in that district are 1:19, the LR of sexual behaviors is about 7:1, and the posterior odds of abuse are 7:19. Therefore about 27% of the reports investigated by Worker A are valid, and the rest are not. Worker A may conclude on the basis of professional experience that sexual behaviors are a very unreliable indicator of sexual abuse.

Worker B deals with the same kind of reports as Worker A but in the Dismal School District. The prior odds of abuse in the Dismal School District are 1:4, the LR of the evidence is again 7:1, and the posterior odds of abuse are 7:4. Therefore about 64% of the reports investigated by Worker B are valid. Worker B may conclude that sexual behaviors are a rather strong indicator of sexual abuse.

If Worker A and Worker B ever work together, they may find themselves vehemently disagreeing about whether or not to validate particular reports that involve sexual behaviors. Their personal experiences will have led them to quite different viewpoints about the proper interpretation of the evidence.

The use of indicators in sexual abuse evaluations is still in a rudimentary stage of development. Badly needed are empirical validation, explicit estimates of indicators' strength, and assessment procedures that take the prior odds of abuse into account.

Like Workers A and B, most of us do not realize how dramatically base rates and prior odds can influence the interpretation of evidence. This lack of understanding can lead to confusion, frustration, and errors. As another example, consider a child protection agency that deals with high-risk children for whom the prior probability of sexual abuse is 70%. The workers in the agency may observe that a substantial number of these children have nightmares, and that over 80% of the children with nightmares have in fact been abused.

If the workers do not understand the importance of base rates and prior odds, they may mistakenly conclude that nightmares are strong evidence of abuse. In fact nightmares probably constitute very weak evidence of abuse, with an LR of perhaps 2.3:1 (Wells, McCann, Adams, Voris, & Ensign, 1995). However, even such weak evidence as nightmares can raise the posterior probability of abuse to 84%, if the prior probability is 70% ($7/3 \times 2.3/1 = 16/3 = 84\%$ probability).

Consider the problems that can arise, however, if the agency then alerts the community: "Nightmares can be a sign of sexual abuse. In our experience, over 80% of children with nightmares have been sexually abused." The agency may be flooded with reports from parents, teachers, and therapists regarding children with nightmares. Most of these new reports will be invalid. Furthermore, the LR of nightmares will change among the children seen by the agency. Because the agency is now seeing a large number of nonabused children with nightmares, the LR of nightmares may change from 2.3:1 to 1:1, or even to 1:2.3. In other words, nightmare reports may actually become evidence that abuse has *not* occurred.

We suspect that many CPS workers will see a connection between the foregoing example and their own experience (see Faller, 1985). Bayes's theorem helps explain what can go wrong when uninformed personal experience guides the interpretation of evidence.

Indicators of Abuse

Controversy currently surrounds the use of indicators to evaluate sexual abuse. Professionals have identified certain child behaviors (e.g., nightmares, bed wetting, sexual behaviors) and case characteristics (e.g., delayed disclosure, concurrent custody dispute) as indicators that a sexual abuse allegation is true or false (e.g. Faller, 1988; Gardner, 1987; Schetky, 1988; Sgroi, Porter, & Blick, 1985). However, other professionals have questioned the use of such indicators (Berliner & Conte, 1993; Finkelhor, 1993; see also Conte, 1992). For example, Berliner and Conte (1993) argue that some proposed indicators lack demonstrated validity (e.g., Gardner, 1987), are common among nonabused children, and should not be considered "determinative" of abuse.

This problem can be analyzed within a framework based on Bayes's theorem. According to this framework, four criteria must be met before an indicator should be used in a sexual abuse evaluation:

1. The indicator must be valid, with an LR greater than 1 (i.e., the indicator behavior must be significantly more frequent among abused than nonabused children).
2. The strength of the indicator (the LR) must be known, at least approximately. If numerical estimates of the LR are unavailable, then common language equivalents (e.g., weak, very strong) may be used, although with some loss of precision.
3. The prior odds of abuse must be known, at least approximately. If precise figures are unavailable (National Resource Center on Child Sexual Abuse, 1993), then reasonable estimates or a range of estimates may be used.
4. The prior odds must be taken into account when conclusions are drawn from the evidence.

According to these criteria, many proposed indicators of sexual abuse are clearly inadequate as a basis for rational decision making, at least at the present time.

First, as pointed out by Berliner and Conte (1993), many indicators are based on clinical impressions only and have not been scientifically validated (e.g. Gardner, 1987; Schetky, 1988; Sgroi et al., 1985).

Second, information regarding the *strength* of indicators is often lacking (e.g. Faller, 1988; Schetky, 1988; Sgroi et al., 1985). Published lists of indicators often indiscriminately lump together evidence that is likely to be strong (e.g., imitation of intercourse), moderate (e.g., sexual play with dolls) and weak (e.g., nightmares, bed wetting).

Third, when such lists are published, there is seldom any acknowledgment that the indicators *by themselves* are insufficient to make determinations of abuse. The subject of prior odds and its relevance to decision making are typically ignored in discussions of indicators.

Thus, the present analysis leads to conclusions similar to those of Berliner and Conte (1993) and Finkelhor (1993). The use of indicators in sexual abuse evaluations is still in a rudimentary stage of development. Badly needed are (a) empirical validation, (b) explicit estimates of indicators' strength, and (c) assessment procedures that take the prior odds of abuse into account.

Despite these problems, there remains some room for optimism. Most important, some proposed indicators of abuse do have demonstrated validity. For example, research has confirmed that some sexual behaviors are much more common among abused than nonabused children, particularly at preschool ages (see reviews by Friedrich, 1993; Kendall-Tackett et al., 1993). Such indicators are potentially useful in evaluations.

It should be added that terms such as *confirming, disconfirming,* and *determinative* are probably best avoided in discussions regarding indicators (Berliner & Conte, 1993; Finkelhor, 1993). The difficulty with such terms is their ambiguity. For example, imagine that a student from the Dismal School District has displayed sexual behaviors. The prior odds of abuse in the district are 1:4, and the LR for sexual behavior is 7:1. The posterior odds of abuse are therefore 7:4, and the posterior probability is 64%.

By the standards of many Child Protective Services, which require only a preponderance of the evidence (probability greater than 50%), the presence of the indicator is sufficient to substantiate abuse and is therefore determinative. However, by the standards of the criminal justice system, which require evidence beyond a reasonable doubt (probability greater than 95%), the evidence is insufficient to prove abuse and thus is not determinative.

Furthermore, if the same evidence (sexual behavior) were present in a case from the Protective School District, where the rate of abuse is 5%, then the posterior probability of abuse would be 27%, which is not determinative, even by the standard of preponderance of the evidence. In fact, a posterior probability of 27% tends to *disconfirm* abuse.

Because terms such as determinative and confirming are ambiguous, their use will inevitably lead to confusion and misunderstandings. Therefore, we suggest that *evidence* be described with the terms from Table 1 (i.e., weak, moderate, strong, very strong) and that *conclusions* be described using the terms from Table 3 (e.g., somewhat more likely than not, very likely).

Techniques of Child Interviewing

Controversy currently surrounds the use of certain techniques in child abuse interviews, particularly with young children (Lamb, 1994; White & Quinn, 1988; see also review by Ceci & Bruck, 1993). Some of these techniques include (a) suggestive or leading questions, (b) repeatedly posing a question that the child has already answered once, (c) flatly contradicting the child, (d) rewarding the child for statements, (e) telling the child that "other people" have already told the interviewer that the child was molested, or (f) making conjectural statements as if they were certain.

The use of such problematic techniques may be defended on two grounds. First, it may be argued that such techniques seldom cause children to make false statements. However, this argument is not well-supported by scientific evidence. On the contrary, research shows that at least some of these techniques do appreciably increase the risk of a false statement, particularly among preschool children (Ceci & Bruck, 1993).

A second argument is that the techniques may be necessary to elicit statements from some children, particularly children who are very young or have been pressured not to talk. For example, research indicates that children are less willing to report sexual abuse if their parents are unsupportive (Lawson & Chaffin, 1992). Perhaps problematic interviewing techniques are necessary to overcome reticence and avoid "false negatives" in sexual abuse evaluations.

What happens if the problematic techniques increase the probability of both true *and* false statements? Bayes's theorem helps to clarify the situation. Consider the following examples:

> *Example 1:* Interviewer X meticulously avoids the problematic techniques mentioned above and relies entirely on open-ended questioning of the child. With this nonaggressive approach, Interviewer X obtains descriptions of sexual abuse from only 20% of the children who have in fact been abused. On the other hand, the number of false statements is very low: only 1% of nonabused children give false descriptions of abuse.
>
> *Example 2:* Eager to avoid false negatives, Interviewer Y vigorously employs the problematic techniques. They work very well: 100% of the abused children give descriptions of having been abused. Unfortunately, so do 20% of the nonabused children.

The numbers in these examples are not based on real data and are introduced only to illustrate a counterintuitive insight: Vigorous attempts to elicit statements can sometimes inadvertently degrade the quality of evidence.

At first glance, Interviewer Y seems to have done much better than Interviewer X. Interviewer Y has obtained 80% more true descriptions of abuse than Interviewer X (100% – 20%), but only 19% more false descriptions (20% – 1%). To the uninstructed eye, Interviewer Y appears to have reaped considerable benefit at very little cost.

However, the LRs tell a different story. A child's statement obtained by Interviewer X has an LR of 20:1 and constitutes strong evidence of abuse. In contrast, a statement obtained by Interviewer Y has an LR of 100:20, or 5:1, and constitutes only weak-to-moderate evidence of abuse. A statement obtained by Interviewer X is probably strong enough to serve as a basis for swift legal action (depending on the prior odds). By contrast, a statement obtained by Interviewer Y is probably too weak to establish that abuse has occurred.

If we assume that the use of problematic techniques increases true reporting, but at the price of increasing false reporting as well, what is a child interviewer to do? One very helpful solution is provided in the *Memorandum of Good Practice* regarding child interviews published by the British government (Home Office, 1992; see also American Professional Society on the Abuse of Children, 1990; Jones & McQuiston, 1988). The Memorandum (Home Office, 1992) suggests that the interviewer begin by using open-ended questions with the child. If open-ended questions do not result in a statement, the interviewer may proceed to closed questions. If closed questions do not result in a statement, the interviewer may proceed to questions that involve a mild level of suggestiveness.

The approach suggested by the Memorandum is consistent with the insights provided by Bayes's theorem. First, an attempt is made to obtain the strongest kind of evidence (a spontaneous statement in response to open-ended questions). If no statement is obtained from the child, then an attempt is made to obtain a somewhat weaker kind of evidence (a statement in response to closed or suggestive questions).

Such a stepwise approach (Yuille, Hunter, Joffe, & Zaparniuk, 1993) guarantees that the strongest statement possible will be obtained from the child. In addition, the approach allows the interviewer to avoid false negatives when the child is withdrawn or reluctant to talk at the beginning of the interview.

Sexual Abuse Allegations in the Context of Custody Disputes

Some professionals have argued that custody disputes encourage false allegations of sexual abuse (e.g., Gardner, 1987). Others have responded that allegations arising in the context of custody disputes are frequently true and should be taken seriously by evaluators (Thoennes & Tjaden, 1990; see also Faller, Corwin, & Olafson, 1993). Bayes's theorem can shed light on this problem.

Research suggests that "in general," about 70% of sexual abuse allegations are reliable (Jones & McGraw, 1987). When allegations arise in the context of custody disputes, the number appears to be closer to 50% (Thoennes & Tjaden, 1990). Within the framework provided by the theorem, therefore, the prior odds of an allegation being reliable are generally about 7:3 (70% vs. 30%). Similarly, the posterior odds (after the evidence of custody dispute has been evaluated) are about 1:1 (50% vs. 50%). By inserting these odds into Bayes's theorem, an approximate LR can be calculated for the evidentiary strength of custody dispute.

$$\text{Prior odds} \times \text{Likelihood ratio} = \text{Posterior odds}$$

$$\frac{7}{3} \times \frac{1}{2.33} = \frac{1}{1}. \qquad \text{(I)}$$

As can be seen, the LR for custody dispute is apparently about 1:2.3 or 1:2, the figure that was mentioned earlier in this article. An LR this small indicates evidence that is weak to very weak.

Thoennes and Tjaden (1990) have concluded that sexual abuse allegations arising in the context of custody disputes are as likely to be true as sexual abuse allegations in general. Although the present analysis differs somewhat from that of Thoennes and Tjaden (1990), it leads to virtually the same practical conclusion: It is a serious error to dismiss sexual abuse allegations because they arise in the context of a custody dispute. The evidence of custody dispute is very weak indeed and should have little or no influence on professional decisions regarding the allegations.

BAYES'S THEOREM AS AN AID FOR SEXUAL ABUSE EVALUATORS

The preceding discussion has applied Bayes's theorem to general issues regarding sexual abuse. Although evaluators can also apply the theorem to individual cases of alleged abuse, a cautious approach is advisable. In any evaluation, complications and uncertainties arise that have not been fully treated in the present article. First, the LRs of evidence are seldom known with any precision. For instance, for instructional purposes the present article has repeatedly referred to the findings of Friedrich et al. (1992) regarding children's sexual behaviors. However, LRs based on those findings may change in the light of future research or may not apply to all children, regardless of age, gender, or ethnic group.

Base rates provide a second source of uncertainty. For example, the present article has several times referred to the finding of Jones and McGraw (1987) that 70% of sexual abuse allegations in a particular child protective agency were reliable. However, that finding came from one agency almost a decade ago and was based on subjective judgments. Thus, the figure of 70% cannot be considered definitive or universal.

Third, base rates and LRs can shift dramatically between groups or, over time, within the same group. As discussed above, for example, if all cases of nightmares are referred routinely to a CPS for evaluation, then the base rate of valid cases will decrease, as will the LR of nightmares. The same problem applies to other kinds of evidence.

Fourth, complications arise when a case involves more than one piece of evidence. For example, the evaluator may have to consider (a) physical evidence, (b) the child's statement, (c) statements by witnesses, (d) the alleged perpetrator's history, and (e) the context in which allegations arose. Although Bayes's theorem can be applied when there are multiple pieces of evidence, the procedure is too complicated for treatment in this article.

Given the uncertainties just described, Bayes's theorem cannot provide a scientifically sure method for sexual abuse evaluations. In most real cases, the relevant LRs and base rates can be estimated only roughly. An evaluator must usually improvise confidence limits, such as "Among cases referred to me for evaluation, I think the base rate of abuse is probably somewhere between 40% and 60%," or "According to the research literature, bed wetting is probably weak evidence of abuse, with an LR between 1:1 and 3:1."

The disappointing truth is that the application of Bayes's theorem to a particular sexual abuse case usually yields inexact results. Both the input and output of the theorem lack precision. On the other hand, if used cautiously as a self-check or aid, the theorem can frequently clarify issues in a particular case. In our experience, unexpected insights arise once an evaluator starts routinely asking questions such as how strong is this piece of evidence or what is the prior probability of abuse for this child.

The factual case study presented at the beginning of this article illustrates the potential helpfulness of the theorem. As will be recalled, sperm was found in the rectum of 5-year-old Billy B. following visitation with his father, Alfonso. A court-appointed evaluator, Dr. Alcoa, concluded that Billy's mother had placed the sperm in his bottom in order to substantiate allegations of sexual abuse against her ex-husband.

How might Dr. Alcoa have gone about evaluating this case in the light of Bayes's theorem? We may begin by reviewing the relevant facts. First, according to Dr. Alcoa's testimony, she believed that only two persons could have placed the sperm on Billy's bottom: his father or his mother. Second, Dr. Alcoa stated that there was only one piece of evidence that the mother had planted the sperm: Billy's statement to Detective Morales of the police department. Dr. Alcoa described the statement as follows:

> Billy reported to Detective Morales that his mother put ointment in his bottom prior to taking him to the hospital. He related to Detective Morales that his mother made him take off his pants, then she took off his underwear, and she put some ointment in his bottom. Billy gave conflicting information with regard to whether this occurred prior to his visit with his father or afterward. However, he did clearly state that his mother put the ointment in his bottom on the same day she took him to be examined for alleged sexual assault at the hospital.

Aside from this statement, Dr. Alcoa knew of no evidence that implicated the mother. Dr. Alcoa also

$$\frac{1}{99} \times \frac{55}{1} = \frac{55}{99} = \text{about } 36\% \text{ probability mother was perpetrator.} \qquad \text{(J)}$$

agreed that Detective Morales had obtained the statement from Billy by using leading questions.

Dr. Alcoa faced a question that was eminently amenable to analysis using Bayes's theorem: Who was more likely to have placed the sperm on Billy's bottom, his mother or his father? To answer such a question, the prior odds must first be calculated. In general, when sperm is found in a child's body, what is the probability that the perpetrator is female rather than male? No precise statistics are available, but 1% is probably an upper-end estimate. The prior odds of a female versus a male perpetrator are certainly no greater than 1:99.

Second, the LR must be calculated for Billy's statement regarding the ointment. One way to estimate an LR would be to analyze the statement for qualities such as consistency, detail, and so on. Such an analysis (undertaken later by a second psychologist) found that although Billy was intellectually and verbally capable of giving detailed descriptions of his experience, his statement regarding the ointment lacked detail, was logically incoherent, contained several major internal contradictions besides those regarding the time of day, and was made in response to leading and coercive questions.

Based on the second psychologist's analysis, the evidence against the mother would probably have to be considered weak (LR = 3:1) or moderate (LR = 7:1) at most. Few evaluators would regard the statement as strong (LR = 20:1) or even very strong (LR = 55:1). In a sense, however, the exact strength of the evidence is not very important in this instance. No matter which of these various LRs is selected, Bayes's theorem leads to the same conclusion: The sperm in Billy's rectum was probably deposited there by his father, not his mother.

This conclusion is almost unavoidable because the prior odds in the mother's favor were extreme (1:99). When the prior odds are extreme in one direction, even very strong evidence in the opposite direction may count for little (see Meehl & Rosen, 1955). For example, let us assume, very liberally, that Billy's statement regarding the ointment constituted very strong evidence against the mother (LR = 55:1). Using Bayes's theorem, the posterior odds that the mother was the perpetrator can be calculated in equation J above.

This example has deliberately stacked the deck against the mother by assuming that the prior odds in her favor were only 1:99 (which is doubtful) and that the evidence against her was very strong (which is even more doubtful). Even with these assumptions, the probability that she was the perpetrator was only about 36%, compared with a 64% probability for the father.

SUMMARY AND CONCLUSIONS

In this article, Bayes's theorem has been introduced to professionals who work in the field of child sexual abuse. The theorem illuminates general issues and can aid evaluators who assess individual cases. Decisions regarding sexual abuse can have the gravest impact on children's lives. It is important, therefore, that evaluations be based on reliable information and careful reasoning. An irrational, careless, or ill-founded decision by an evaluator can lead to tragic results. It is worth noting that the opinions offered by Dr. Alcoa in the case study presented above were accepted without disagreement by the local Child Protection Services and a judge. Based on Dr. Alcoa's evaluation of the case, Billy was eventually removed from his mother and placed into the permanent custody of his father.

REFERENCES

American Professional Society on the Abuse of Children. (1990). *Guidelines for psychosocial evaluation of suspected sexual abuse in young children.* Chicago: Author.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society, 53,* 370.

Berliner, L., & Conte, J. R. (1993). Sexual abuse evaluations: Conceptual and empirical obstacles. *Child Abuse and Neglect, 17,* 111-125.

Ceci, S. J., & Bruck, M. (1993). Suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin, 113,* 403-439.

Conte, J. R. (1992). Has this child been sexually abused? *Criminal Justice and Behavior, 19,* 54-73.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243,* 1668-1674.

Faller, K. C. (1985). Unanticipated problems in the United States child protection system. *Child Abuse and Neglect, 9,* 63-69.

Faller, K. C. (1988). Criteria for judging the credibility of children's statements about their sexual abuse. *Child Welfare, 67,* 389-401.

Faller, K. C., Corwin, D. L., & Olafson, E. (1993). Research on false allegations of sexual abuse in divorce. *APSAC Advisor, 6*(3), 1, 7-10.

Finkelhor, D. (1993). Epidemiological factors in the clinical identification of child sexual abuse. *Child Abuse and Neglect, 17,* 67-70.

Fisher, R. A. (1959). *Statistical methods and scientific inference.* London: Oliver & Boyd.

Friedrich, W. N. (1993). Sexual victimization and sexual behavior in children: A review of recent literature. *Child Abuse and Neglect, 17,* 59-66.

Friedrich, W. N., Grambsch, P., Damon, L., Hewitt, S. K., Koverola, C., Lang, R. A., Wolfe, V., & Broughton, D. (1992). Child Sexual Behavior Inventory: Normative and clinical comparisons. *Psychological Assessment, 4,* 303-311.

Gardner, R. A. (1987). The Sexual Abuse Legitimacy Scale. In R. A. Gardner (Ed.), *The parental alienation syndrome and the differentia-*

tion between fabricated and genuine child sex abuse (pp. 171-210). Cresskill, NJ: Creative Therapeutics.

Goodman, S. N., & Royall, R. (1988). Evidence and scientific research. *American Journal of Public Health. 78*, 1568-1574.

Home Office. (1992). *Memorandum of good practice on video recorded interviews with child witnesses for criminal proceedings*. London: HMSO.

Jones, D.P.H., & McGraw, J. M. (1987). Reliable and fictitious accounts of sexual abuse in children. *Journal of Interpersonal Violence, 2*, 27-45.

Jones, D.P.H., & McQuiston, M. G. (1988). *Interviewing the sexually abused child*. London: Gaskell.

Kendall-Tackett, K. A., Williams, L. M., & Finkelhor, D. (1993). Impact of sexual abuse on children: A review and synthesis of recent empirical studies. *Psychological Bulletin, 113*, 164-180.

Lamb, M. E. (1994). The investigation of child sexual abuse: An interdisciplinary consensus statement. *Child Abuse and Neglect, 18*, 1021-1028.

Lawson, L., & Chaffin, M. (1992). False negatives in sexual abuse disclosure interviews. *Journal of Interpersonal Violence, 7*, 532-542.

Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin, 52*, 194-216.

National Resource Center on Child Sexual Abuse. (1993). Statistical base still unfolding. *NRCCSA News, 1*, 5, 8.

Salter, A. C. (1988). *Treating child sex offenders and victims*. Newbury Park, CA: Sage.

Schetky, D. H. (1988). The clinical evaluation of child sexual abuse. In D. H. Schetky & A. H. Green (Eds.), *Child sexual abuse* (pp. 57-81). New York: Brunner/Mazel.

Sgroi, S. M., Porter, F. S., & Blick, L. C. (1985). Validation of child sexual abuse. In S. M. Sgroi (Ed.), *Handbook of clinical intervention in child sexual abuse* (pp. 39-79). Lexington, MA: Lexington Books.

Thoennes, N., & Tjaden, P. G. (1990). The extent, nature, and validity of sexual abuse allegations in custody/visitation disputes. *Child Abuse and Neglect, 14*, 151-163.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty. *Science, 185*, 1124-1131.

Wells, R. D., McCann, J., Adams, J., Voris, J., & Ensign, J. (1995). Emotional, behavioral, and physical symptoms reported by parents of sexually abused, nonabused, and allegedly abused prepubescent females. *Child Abuse and Neglect, 19*, 155-163.

White, S., & Quinn, K. M. (1988). Investigatory independence in child sexual abuse evaluations: Conceptual considerations. *Bulletin of the American Academy of Psychiatry and the Law, 16*, 269-278.

Wood, J. M., & Wright, L. (1995). Evaluation of children's sexual behaviors and incorporation of base rates in judgments of sexual abuse. *Child Abuse and Neglect, 19*, 1263-1273.

Yuille, J. C., Hunter, R., Joffe, R., & Zaparniuk, J. (1993). Interviewing children in sexual abuse cases. In G. S. Goodman & B. L. Bottoms (Eds.), *Child victims, child witnesses* (pp. 95-115). New York: Guilford.

James M. Wood, Department of Psychology, University of Texas at El Paso. *I gratefully acknowledge M. Teresa Nezworski, Lee Sechrest, and Randolph Whitworth for comments on a draft of this article.* *Correspondence and requests for reprints should be sent to James M. Wood, Department of Psychology, University of Texas at El Paso, El Paso TX 79968.*