

Relevance Ratio: Evaluating the Probative Value of Expert Testimony in Child Sexual Abuse Cases

Thomas D. Lyon

Jonathan J. Koehler

Follow this and additional works at: <http://scholarship.law.cornell.edu/clr>

 Part of the [Law Commons](#)

Recommended Citation

Thomas D. Lyon and Jonathan J. Koehler, *Relevance Ratio: Evaluating the Probative Value of Expert Testimony in Child Sexual Abuse Cases*, 82 Cornell L. Rev. 43 (1996)

Available at: <http://scholarship.law.cornell.edu/clr/vol82/iss1/4>

This Article is brought to you for free and open access by the Journals at Scholarship@Cornell Law: A Digital Repository. It has been accepted for inclusion in Cornell Law Review by an authorized administrator of Scholarship@Cornell Law: A Digital Repository. For more information, please contact jmp8@cornell.edu.

THE RELEVANCE RATIO: EVALUATING THE PROBATIVE VALUE OF EXPERT TESTIMONY IN CHILD SEXUAL ABUSE CASES

Thomas D. Lyon† & Jonathan J. Koehler††

I INTRODUCTION

It is hard to overstate the importance of expert testimony in American courtrooms. Much of this testimony concerns scientific matters that are beyond the ken of ordinary experience. In cases where scientific matters play a central role, jurors may give substantial weight to expert testimony or even treat it as dispositive. Standards pertaining to the admissibility of scientific testimony are critical to the outcome in many trials.

In *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, the United States Supreme Court adopted a new standard for the admissibility of expert scientific testimony.¹ Under *Daubert*, courts must independently assess the scientific validity of an expert's methodology. The Court explicitly rejected the *Frye* rule, under which courts based the admissibility of expert testimony entirely on whether the expert's methodology had achieved general acceptance in the relevant scientific community.² Whether a technique is "generally accepted" is now only one consideration in a multi-pronged analysis, which includes the falsifiability of the expert's theory and the scientific technique's rate of error.³

Daubert could not have come at a better time. In recent years, critics have attacked expert testimony in many areas as "junk science"⁴

† Thomas D. Lyon (J.D. 1987, Harvard Law School, Ph.D. 1994, Stanford University, Department of Psychology) is Assistant Professor of Law at the University of Southern California Law School (University Park, Los Angeles, CA 90089-0071, e-mail: tlyon@law.usc.edu).

†† Jonathan J. Koehler (Ph.D. 1989, University of Chicago, Department of Behavioral Sciences) is Associate Professor of Behavioral Decision Making in the Business and Law Schools at the University of Texas at Austin (CBA 5.202, Austin, TX 78712, e-mail: koehler@mail.utexas.edu).

¹ 509 U.S. 579 (1993).

² *Frye v. United States*, 293 F.2d 1013 (D.C. Cir. 1923).

³ See *Daubert*, 509 U.S. at 593-94. The inquiry also includes a consideration of whether the theory has been subject to peer review and publication, and whether standards exist by which to evaluate the proffered technique. *Id.*

⁴ See, e.g., PETER W. HUBER, *GALILEO'S REVENGE: JUNK SCIENCE IN THE COURTROOM* (1991).

or, less colorfully, as lacking rigorous empirical support.⁵ Some of the harshest criticism has been reserved for expert psychological and psychiatric testimony in cases involving allegations of child sexual abuse.⁶ Some commentators have argued that much of what passes for scientific knowledge in this field is more impressionistic and anecdotal than scientific.⁷ *Daubert* mandates that trial courts take these criticisms seriously. Courts must not only demand that experts provide empirical support for their claims, but they must also evaluate the scientific merits of the available data before making admissibility decisions.⁸

Whether trial judges are capable of performing like "amateur scientists"⁹ is questionable. What makes a test falsifiable?¹⁰ What is

⁵ See generally Charles P. Ewing, "Dr. Death" and the Case for an Ethical Ban on Psychiatric and Psychological Predictions of Dangerousness in Capital Sentencing Proceedings, 8 AM. J.L. & MED. 407, 417-18 (1983) (arguing that predictions of dangerousness "cannot be said to be founded on a scientific basis"); Michael J. Saks & Jonathan J. Koehler, *What DNA "Fingerprinting" Can Teach the Law About the Rest of Forensic Science*, 13 CARDOZO L. REV. 361, 372 (1991) ("[M]ost forensic sciences, including DNA typing, rely on assumptions that have not yet been verified by empirical testing."); David L. Faigman, Note, *The Battered Woman Syndrome and Self-defense: A Legal and Empirical Dissent*, 72 VA. L. REV. 619 (1986) (criticizing the quality of research on battered woman syndrome and arguing that courts should not admit expert testimony based on this research).

⁶ See generally Ronald J. Allen & Joseph S. Miller, *The Expert as Educator: Enhancing the Rationality of Verdicts in Child Sex Abuse Prosecutions*, 1 PSYCHOL., PUB. POL'Y, & L. 323 (1995) (promoting expert testimony only as a means of educating jurors); Lisa R. Askowitz & Michael H. Graham, *The Reliability of Expert Psychological Testimony in Child Sexual Abuse Prosecutions*, 15 CARDOZO L. REV. 2027 (1994) (urging courts to take a more active role in scrutinizing expert psychological testimony before admitting it in child abuse prosecutions); Robert J. Levy, *Using "Scientific" Testimony to Prove Child Sexual Abuse*, 23 FAM. L.Q. 383 (1989) (concluding that mental health professionals' testimony about evaluative techniques that are unscientific and deceptive jeopardizes the fairness of factual findings); Mary Ann Mason, *The Child Sex Abuse Syndrome: The Other Major Issue in State of New Jersey v. Margaret Kelly Michaels*, 1 PSYCHOL., PUB. POL'Y, & L. 399 (1995) (supporting a closer examination of scientific reliability in child abuse cases); David McCord, *Expert Psychological Testimony About Child Complainants in Sexual Abuse Prosecutions: A Foray into the Admissibility of Novel Psychological Evidence*, 77 J. CRIM. L. & CRIMINOLOGY 1, 6 (1986) (noting "a woeful lack of serious review of the existing behavioral scientific research [on child abuse] by the courts"); Andrew Cohen, Note, *The Unreliability of Expert Testimony on the Typical Characteristics of Sexual Abuse Victims*, 74 GEO. L.J. 429 (1985) (advocating the admission of expert testimony on behalf of sexual abuse victims only in unusual cases because such testimony often misleads a jury).

⁷ See, e.g., Mason, *supra* note 6, at 402 (asserting that current research does not support clinicians' claims that there are clear behavioral indicators of sexual abuse).

⁸ *Daubert*, 509 U.S. at 592-93 (holding that Rule 104(a) requires federal judges to make a "preliminary assessment of whether the reasoning or methodology underlying the testimony is scientifically valid and of whether that reasoning or methodology properly can be applied to the facts in issue").

⁹ Chief Justice Rehnquist used this term in *Daubert*, 509 U.S. at 601 (Rehnquist, C.J., concurring in part and dissenting in part).

¹⁰ As Chief Justice Rehnquist complained, "I defer to no one in my confidence in federal judges; but I am at a loss to know what is meant when it is said that the scientific status of a theory depends on its 'falsifiability,' and I suspect some of them will be, too." *Id.* at 600 (Rehnquist, C.J., concurring in part and dissenting in part).

meant by a technique's "rate of error"?¹¹ How much error is tolerable? The tests outlined in *Daubert* are worded as "general observations,"¹² rather than specific tools for assessing the scientific validity or the probative value of proffered testimony.¹³

Our paper addresses this problem. We feel strongly that the admissibility of scientific expert testimony should turn on its scientific merit. However, we also sympathize with the difficulties that trial courts face when trying to think like scientists. As a first step toward solving this dilemma, courts can make major advances in the treatment of proffered scientific testimony by recasting certain traditional legal concepts in scientifically appropriate ways.

One important legal concept is relevance. All evidence must be relevant to be admissible.¹⁴ This Article offers a rigorous definition of relevance that courts can use to evaluate the probative value of proffered scientific testimony. We introduce what we call the relevance ratio—a simple formula that is based on the legal definition of relevant evidence and on probability theory. Even judges and others who have little background in science or mathematics may use the relevance ratio profitably. We apply this ratio in the context of expert testimony in child sexual abuse cases. In doing so, we examine how the admissibility decision would be better informed if courts used the relevance ratio. The ratio is equally useful for assessing the relevance of expert testimony in other fields.

Part II defines the relevance ratio and explains its relation to probabilistic reasoning. Part III uses the ratio to explore the ways in which physicians have misused the term "consistent with sexual abuse" in child abuse cases. Part IV considers whether symptoms "consistent with sexual abuse," although irrelevant for the purpose of proving abuse, may nevertheless be admissible for rebutting the assertion that abuse did not occur. Part V uses the relevance ratio to show that symptoms that are common among abused children may nonetheless have little if any relevance for proving that abuse occurred. Part VI discusses the relevance of clusters of symptoms and reveals ways in which clusters are likely to be less relevant than often believed. Part VII outlines the methodological limitations of existing research on the

¹¹ See, e.g., *infra* notes 62-69 and accompanying text (misconceptions about the "rate of error").

¹² *Daubert*, 509 U.S. at 593.

¹³ See Margaret A. Berger, *Procedural Paradigms for Applying the Daubert Test*, 78 MINN. L. REV. 1345, 1345 (1994).

¹⁴ In this paper, we use "relevance" to refer to logical relevance. Although *Daubert* discusses relevance, it does so in terms of "fit," which is analogous to materiality and concerns the extent to which the expert testimony is "tied to the facts of the case." *Daubert*, 509 U.S. at 591 (quoting *United States v. Downing*, 753 F.2d 1224, 1242 (3rd Cir. 1985)). We do not address the issue of "fit" in this Article.

symptoms of child sexual abuse and explains why they may lead to poor estimates of the relevance ratio. Part VIII uses the relevance ratio to demonstrate that probative asymmetries exist between the presence and absence of various symptoms. Part IX discusses the significance of the existence of symptoms among nonabused children for understanding the significance of those symptoms among abused children.

II

THE RELEVANCE RATIO

Federal Rule of Evidence 401 defines "relevant evidence" as evidence that has "any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence."¹⁵ In a child abuse case, evidence that a child suffered a particular symptom is relevant for proving abuse if the presence of the symptom increases the chance that abuse actually occurred. One can determine whether the presence of symptoms increases the chance that abuse occurred by considering two proportions: the proportion of abuse cases in which the symptom occurs, and the proportion of nonabuse cases in which the symptom occurs. If the proportion of abuse cases exhibiting the symptom is greater than the proportion of nonabuse cases exhibiting the symptom, then the symptom is relevant for proving that abuse occurred.¹⁶ We refer to this ratio of proportions as the "relevance ratio" and offer this ratio as the appropriate standard against which to measure the relevance of expert testimony.¹⁷

¹⁵ See FED. R. EVID. 401; see also CHRISTOPHER B. MUELLER & LAIRD C. KIRKPATRICK, *MODERN EVIDENCE* § 1.2, at 4 n.2 (1995) (noting that as of 1994, 35 states have adopted rules of evidence similar to the Federal Rules).

¹⁶ Cf. Richard O. Lempert, *Modeling Relevance*, 75 MICH. L. REV. 1021, 1026 (1977) ("[E]vidence is logically relevant only when the probability of finding that evidence given the truth of some hypothesis at issue in the case differs from the probability of finding the same evidence given the falsity of the hypothesis at issue"). The logarithm of the relevance ratio (i.e., Bayes's "likelihood ratio") provides an equivalent measure of relevance. This method is advantageous, because it equates the relevance of evidence offered by both the plaintiff and the defendant. See David H. Kaye, Comment, *Quantifying Probative Value*, 66 B.U. L. REV. 761, 765 (1986) ("[E]vidence is relevant as that term is used in Rule 401 if its log-likelihood ratio is not zero."); see also Ward Edwards, Comment, 66 B.U. L. REV. 623, 626 (1986) (noting a log-likelihood ratio's properties of symmetry and additivity, qualities that other measures lack). Although we agree with Kaye and Edwards that the log-likelihood ratio has some desirable mathematical properties, we believe that the simplicity and intuitive appeal of the relevance ratio make it the better candidate for heuristic use by judges. Moreover, as Kaye points out, one can maintain the mathematical symmetry between the relevance ratio for the plaintiff's evidence and the relevance ratio for the defendant's evidence by simply inverting the relevance ratio when considering the defendant's evidence. Kaye, *supra*, at 763 n.12.

¹⁷ See *infra* fig.1.

For example, suppose a prosecutor wishes to introduce expert psychological testimony that a particular child suffered from low self-esteem to support his contention that the child was abused. Is low self-esteem relevant for proving abuse? Even if data are available which indicate that many abused children suffer from low self-esteem,¹⁸ it may also be true that many nonabused children suffer from low self-esteem. Neither fact in isolation indicates whether low self-esteem is relevant for proving abuse. However, a ratio of the relevant proportions answers this question. If a higher percentage of abused children exhibit low self-esteem than nonabused children, then this symptom is relevant for proving that abuse has occurred. Many studies have examined self-esteem in these two populations and differences have not been detected consistently.¹⁹ Thus, the relevance ratio for low self-esteem appears to be near unity, and one may conclude that this symptom fails a relevance test.

This simple example illustrates the intuition that a symptom is less probative if one commonly observes the symptom among nonabused children. However, as will be discussed shortly, conclusions reached through intuitive appraisals often differ from those reached using a relevance ratio analysis. In our view, *a relevance ratio analysis is the most efficient way to think about evidentiary relevance.*²⁰ In cases of alleged child abuse, one cannot establish the relevance of a symptom or cluster of symptoms merely by confirming that the symptom commonly occurs in abused children. Likewise, dismissing a proffered symptom because the symptom appears in some nonabused children is inappropriate. Relevance is best determined by comparing the relevant proportion in both the abused and nonabused population.

¹⁸ See Angela Browne & David Finkelhor, *Initial and Long-term Effects: A Review of the Research*, in A SOURCEBOOK ON CHILD SEXUAL ABUSE 143, 150 (David Finkelhor et al. eds., 1986) (citing a study in which 58% of sexual abuse victims expressed feelings of inferiority or lack of self-worth).

¹⁹ See Kathleen A. Kendall-Tackett et al., *Impact of Sexual Abuse on Children: A Review and Synthesis of Recent Empirical Studies*, 113 PSYCHOL. BULL. 164, 165 (1993) (noting that half of the studies comparing self-esteem between abused and nonabused children found no difference).

²⁰ But see Richard D. Friedman, *A Close Look at Probative Value*, 66 B.U. L. REV. 733 (1986) [hereinafter Friedman, *A Close Look*]; Richard D. Friedman, *Conditional Probative Value: Neoclassicism Without Myth*, 93 MICH. L. REV. 439, 456-57 (1994). Friedman proposes a measure of probative value that examines the difference between one's beliefs before and after the introduction of evidence. Friedman, *A Close Look*, *supra*, at 738. David Kaye criticizes Friedman's measure for including the strength of one's pre-existing belief in the determinations of the relevance of evidence subsequently received. Kaye, *supra* note 16, at 765-66. But Kaye admits, "There may be more than one right answer to the questions of measuring the probative value of evidence." *Id.* at 766. We agree, but maintain that proposals relying on pre-existing beliefs are questionable for reasons outlined in Kaye, *supra* note 16, and Edwards, *supra* note 16.

Readers familiar with the literature on probabilistic reasoning in the law may recognize the relevance ratio as the likelihood ratio term from Bayes's Theorem. Bayes's Theorem provides a method for updating probabilistic beliefs in the face of new evidence.²¹ It combines a likelihood ratio (which captures the diagnostic value of new evidence) with a prior odds ratio (which captures one's initial beliefs about the hypothesis) to form a posterior odds ratio (post-evidentiary belief about the hypothesis).²²

The applicability of Bayes's Theorem to legal decision-making has been a source of controversy for at least twenty-five years.²³ But,

²¹ MICHAEL O. FINKELSTEIN & BRUCE LEVIN, *STATISTICS FOR LAWYERS* 92-94 (1990).

²² *See id.* The odds form of Bayes's Theorem provides a normatively justifiable mechanism for updating one's beliefs in the face of new evidence. For some Hypothesis H and Evidence E,

$$\frac{P(H|E)}{P(H)} = \frac{P(H)}{P(H)} \times \frac{P(E|H)}{P(E|H)}$$

In other words, the posterior odds ratio

$$\frac{P(H|E)}{P(H)}$$

is the product of the prior odds ratio

$$\frac{P(H)}{P(H)}$$

and the likelihood ratio

$$\frac{P(E|H)}{P(E|H)}$$

See David H. Kaye, *Introduction: What is Bayesianism?*, in *PROBABILITY AND INFERENCE IN THE LAW OF EVIDENCE* 1, 9 (Peter Tillers & Eric D. Green eds., 1988).

²³ *See, e.g.*, Jonathan J. Koehler, *Probabilities in the Courtroom: An Evaluation of the Objections and Policies*, in *HANDBOOK OF PSYCHOLOGY AND LAW* 167 (D.K. Kagehiro & W.S. Laufer eds., 1992); Jonathan J. Koehler, *The Normative Status of Base Rates at Trial*, in *INDIVIDUAL AND GROUP DECISION MAKING* 137 (N. John Castellan, Jr. ed., 1993); Ronald J. Allen, *On the Significance of Batting Averages and Strikeout Totals: A Clarification of the "Naked Statistical Evidence" Debate, the Meaning of "Evidence," and the Requirement of Proof Beyond a Reasonable Doubt*, 65 *TUL. L. REV.* 1093 (1991); Lea Brilmayer & Lewis Kornbauser, *Review: Quantitative Methods and Legal Decisions*, 46 *U. CHI. L. REV.* 116 (1978); Craig R. Callen, *Adjudication and the Appearance of Statistical Evidence*, 65 *TUL. L. REV.* 457 (1991); Craig R. Callen, *Cognitive Science and the Sufficiency of "Sufficiency of the Evidence" Tests*, 65 *TUL. L. REV.* 1113 (1991); Craig R. Callen, *Notes on a Grand Illusion: Some Limits on the Use of Bayesian Theory in Evidence Law*, 57 *IND. L.J.* 1 (1982); Michael O. Finkelstein & William B. Fairley, *A Bayesian Approach to Identification Evidence*, 83 *HARV. L. REV.* 489 (1970); Randolph N. Jonakait, *When Blood is Their Argument: Probabilities in Criminal Cases, Genetic Markers, and, Once Again, Bayes' Theorem*, 1983 *U. ILL. L. REV.* 369; John Kaplan, *Decision Theory and the Factfinding Process*, 20 *STAN. L. REV.* 1065 (1968); David Kaye, *The Paradox of the Gatecrasher and Other Stories*, 1979 *ARIZ. ST. L.J.* 101; D.H. Kaye, *Apples and Oranges: Confidence Coefficients and the Burden of Persuasion*, 73 *CORNELL L. REV.* 54 (1987); Jonathan J. Koehler, *The Probity/Policy Distinction in the Statistical Evidence Debate*, 66 *TUL. L. REV.* 141 (1991); Jonathan J. Koehler & Daniel N. Shaviro, *Veridical Verdicts: Increasing Verdict Accuracy Through the Use of Overtly Probabilistic Evidence and Methods*, 75 *CORNELL L. REV.* 247 (1990); Lempert, *supra* note 16; Charles Nesson, *The Evidence or the Event? On Judicial Proof and the Acceptability of Verdicts*, 98 *HARV. L. REV.* 1357 (1985);

whereas much of the debate has concerned the feasibility of identifying prior probabilities, our analysis focuses on the likelihood ratio. We rely on the research on probative value that points to the likelihood ratio as the most reasonable way of representing relevance²⁴ and maintain that judges should use that function when deciding whether to admit scientific evidence. Professor David Kaye is correct that a likelihood-based approach to relevance or probative value "can be embraced without a commitment to Bayesian reasoning."²⁵ We pursue this perspective vigorously. A consideration of whether Bayes's Theorem should be introduced at trial or whether the trier of fact should be encouraged to use probabilistic combination rules is, therefore, unnecessary. Moreover, to free the likelihood-based approach from its Bayesian baggage and to make the relevance concept more intuitive and less linguistically confusing, we introduce the term "relevance ratio" to replace the term "likelihood ratio." We also dispense with much of the mathematical notation which has discouraged even quantitatively minded jurists from applying the relevance ratio in a wider range of cases.²⁶

Relevance ratio reasoning is also unusual in the testimony provided by experts in psychology and medicine in cases of alleged child abuse. As a result, the terms experts use to describe the frequency or severity of symptoms often obscure the relevance of those symptoms. Moreover, a failure to appreciate the need to consider the frequency

Charles R. Nesson, *Reasonable Doubt and Permissive Inferences: The Value of Complexity*, 92 HARV. L. REV. 1187 (1979); Michael J. Saks & R.F. Kidd, *Human Information Processing and Adjudication: Trial by Heuristics*, 15 L. & SOC'Y REV. 123 (1980-81); Daniel Shaviro, *A Response to Professor Allen*, 65 TUL. L. REV. 1111 (1991); Daniel Shaviro, *A Response to Professor Callen*, 65 TUL. L. REV. 499 (1991); Daniel Shaviro, *Statistical-Probability Evidence and the Appearance of Justice*, 103 HARV. L. REV. 530 (1989); Symposium, *Decision and Inference in Litigation*, 13 CARDOZO L. REV. 253 (1991); Symposium, *Probability and Inference in the Law of Evidence*, 66 B.U. L. REV. 377 (1986); Lawrence H. Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329 (1971); Richard W. Wright, *Causation, Responsibility, Risk, Probability, Naked Statistics, and Proof: Pruning the Bramble Bush by Clarifying the Concepts*, 73 IOWA L. REV. 1001 (1988).

²⁴ See, e.g., Kaye, *supra* note 16; D.H. Kaye, *The Relevance of 'Matching' DNA: Is the Window Half Open or Half Shut?*, 85 J. CRIM. L. & CRIMINOLOGY 676, 683-86 (1995) [hereinafter Kaye, *The Relevance of 'Matching' DNA*]; Lempert, *supra* note 16.

²⁵ Kaye, *The Relevance of 'Matching' DNA*, *supra* note 24, at 678 n.11; see also *United States v. Shonubi*, 895 F. Supp. 460, 485 (E.D.N.Y. 1995) ("It is possible to accept the validity of Bayes Theorem as a guide to . . . decision-making, without attempting the complex calculations literal application . . . would require."), *vacated*, No. 116, 95-1249, 1997 WL 2540 (2d Cir. Jan. 6, 1997); Richard Lempert, *The New Evidence Scholarship: Analyzing the Process of Proof*, 66 B.U. L. REV. 439, 446 (1986) ("Bayes' Theorem may be useful as a heuristic device.").

²⁶ Judge Jack Weinstein (United States District Court for the Eastern District of New York) and Judge Frank Easterbrook (United States Court of Appeals for the Seventh Circuit) are two notable exceptions in an otherwise remarkably math-phobic population. See *Shonubi*, 895 F. Supp. at 460; *Branion v. Gramly*, 855 F.2d 1256 (7th Cir. 1988), *cert. denied*, 490 U.S. 1008.

of symptoms among nonabused children has led many courts and commentators to misinterpret the significance of expert testimony in abuse cases. In what follows, we identify the errors in reasoning attributable to this misunderstanding and explain how an awareness of the relevance ratio can lead to better judgments.

FIGURE 1

The Relevance Ratio

$$\frac{\textit{Proportion of Abused Who Show Symptoms}}{\textit{Proportion of Nonabused Who Show Symptoms}}$$

The relevance ratio measures the relevance or probative value of evidence. In cases of alleged abuse, symptoms associated with ratios greater than one are relevant for proving abuse because they make abuse more likely than it was prior to the introduction of the symptom evidence. Symptoms associated with ratios at or near one are not relevant for proving abuse because they do not change our judgment about the likelihood that abuse occurred.

III

DOES "CONSISTENT WITH" MEAN RELEVANT?

Physicians frequently testify that a child's physical condition is "consistent with" sexual abuse.²⁷ The Eighth Circuit Court of Appeals has asserted that "consistent with" is the "customary cautious professional jargon" for causation.²⁸ Although some medical experts have recommended that the term "consistent with" sexual abuse be limited to cases in which the physical evidence is relevant for proving that abuse occurred,²⁹ a child's physical condition is often described as "consistent with" abuse when the condition is merely concomitant with abuse, but of questionable relevance.³⁰

²⁷ See *infra* notes 29-33. Experts also use the "consistent with" language when describing the behavior of children alleged to have been sexually abused. See, e.g., *State v. Townsend*, 635 So. 2d 949, 958 (Fla. 1994) ("[A] medical expert witness may testify as to whether, in the expert's opinion, the behavior of a child is consistent with the behavior of a child who has been sexually abused.").

²⁸ *United States v. DeNoyer*, 811 F.2d 436, 438 n.3 (8th Cir. 1987).

²⁹ See, e.g., *State v. Gribble*, 804 P.2d 634, 637 (Wash. Ct. App. 1991) (noting that a doctor testified that an attenuated hymen and enlarged vaginal opening were "consistent with," but not absolutely diagnostic of sexual abuse); Jan Bays & David Chadwick, *Medical Diagnosis of the Sexually Abused Child*, 17 CHILD ABUSE & NEGLECT 91, 106 (1993) (listing conditions as "consistent with" sexual abuse that are evidence of abuse but not "diagnostic" of abuse).

³⁰ Doctors will testify that a condition is "consistent with" sexual abuse and also "consistent with" other causes, but not make clear whether the condition is more common among abused children or nonabused children. See, e.g., *State v. Hollywood*, 680 P.2d 655, 657 (Or. Ct. App. 1984) (referencing testimony that an "inflamed and irritated vagina and

A relevance ratio analysis, however, reveals that a condition that is "consistent with" abuse is relevant for proving that abuse occurred only when the condition occurs more frequently among abused children than among nonabused children. As it is commonly used, the "consistent with" terminology does not require consideration of the frequency with which the condition occurs among nonabused children. Typically, the "consistent with" terminology is merely an observation that *at least some* abused children exhibit the condition. Thus, "consistent with" testimony informs a factfinder that the numerator of the relevance ratio is nonzero, but says nothing about the denominator. The numerator must be compared with the denominator, however, for the relevance of the condition to be fully understood.

Physicians testifying in sexual abuse cases have even labeled normal physical findings as "consistent with" sexual abuse. In *Barnes v. United States*, a doctor testified that the lack of physical signs "was consistent with nothing happening at all and with the child's [abuse] story."³¹ In *Randall v. State*, a doctor testified that he found "some opening of the hymenal ring, which could be normal for an 11-year-old child."³² He went on to say, however, that this potentially normal finding was "consistent with" abuse, because "some degree of penetration could occur without totally perforating the hymen."³³ Thus, although normal physical findings are not diagnostic of sexual abuse, physicians often report that they are "consistent with" abuse. Physicians reach this misleading conclusion because abuse often leaves no

outer vaginal lips" in a 4-year-old may be "consistent with sexual molestation" but that it is also consistent with infection); In re Michelle L., 189 A.D.2d 998, 1000 (N.Y. App. Div. 1993) ("[A] pediatrician who examined the girls testified that both girls exhibited redness in their labia and vulva regions, which was consistent with sexual abuse."); *People v. Haun*, 581 N.E.2d 864, 868 (Ill. App. Ct. 1991) (noting that a doctor stated that the "fact she could easily insert her index finger into the victim's vagina . . . [was] consistent with sexual abuse," but "would have also been consistent with the practice of masturbation"); In re Nassau County Department of Social Services, 176 A.D.2d 881, 882 (N.Y. App. Div. 1991) (discussing a witness who explained that a finding of "superficial vaginal injury" [was] "consistent with the possibility of sexual abuse but . . . [was] also consistent with explanations other than sexual abuse"); *State v. Paster*, 524 A.2d 587, 589 (R.I. 1987) ("Doctor Ettefagh's examination revealed anal fissures and what appeared to be an enlarged vaginal opening. Doctor Ettefagh testified that these findings are consistent with sexual abuse, but that they could also result from other causes."). As noted below, redness of the genitalia (erythema) is as common among nonabused children as among abused children. See *infra* note 36.

³¹ 600 A.2d 821, 823 (D.C. 1991).

³² 428 S.E.2d 616, 620 (Ga. Ct. App. 1993).

³³ *Id.*

physical signs³⁴ and because variations in the genitalia are often the product of natural occurrences.³⁵

Studies of sexual abuse that include a control sample (i.e., a group of children who have not been abused) enable professionals to estimate both components of the relevance ratio and thereby estimate the relevance of various physical symptoms. This research shows that some physical conditions that were once widely believed to be indicative of sexual abuse are so common in nonabused children that they have only limited relevance as proof of abuse. Redness of the genitalia (erythema) appears to be as common among nonabused children as among abused children.³⁶ Hymenal diameters among nonabused girls range widely, suggesting that a slightly enlarged vaginal opening is not relevant for proving abuse.³⁷ Other physical findings, once grounds for suspicion, are now considered normal.³⁸ As the relevance

³⁴ See, e.g., Bays & Chadwick, *supra* note 29, at 92 (noting that a review of 21 studies of abused children found that "[n]ormal examinations were reported in 26% to 73% of girls (mean 50%) and 17% to 82% (mean 53%) of boys").

³⁵ See Abbey Berenson et al., *Appearance of the Hymen in Newborns*, 87 PEDIATRICS 458 (1991) (noting that physical examinations of the genitalia of 468 newborn girls showed that lateral and ventral clefts, intravaginal and external ridges, and periurethral bands were common.).

³⁶ See, e.g., Bays & Chadwick, *supra* note 29 (arguing that erythema is common among nonabused girls, and that erythema and perianal erythema are also unlikely to be due to abuse.); S. Jean Emans et al., *Genital Findings in Sexually Abused Symptomatic and Asymptomatic Girls*, 79 PEDIATRICS 778 (1987) (studying 119 abused girls, 127 normal girls with no sexual complaints, and 59 girls with genital complaints unrelated to sexual abuse and concluding that erythema is not significantly more common among abused girls when controlled for race); John McCann et al., *Genital Findings in Prepubertal Girls Selected for Nonabuse: A Descriptive Study*, 86 PEDIATRICS 428, 432 tbl.3 (1990) [hereinafter McCann et al., *Genital Findings*] (finding erythema in 56% of the 91 nonabused girls observed in supine position and in 34% of the 90 nonabused girls observed in the knee-chest position); John McCann et al., *Perianal Findings in Prepubertal Children Selected for Nonabuse: A Descriptive Study*, 13 CHILD ABUSE & NEGLECT 179, 187 tbl.1 (1989) (finding erythema in 41% of the 168 nonabused observed children).

³⁷ See Joyce A. Adams, *Classification of Anogenital Findings in Children with Suspected Sexual Abuse: An Evolving Process*, 6 AMERICAN PROF. SOC'Y ON THE ABUSE OF CHILDREN ADVISOR 11 (1993) (arguing that the hymenal opening may be cause for suspicion if a child has hymenal opening more than two standard deviations above the mean for children of the same age and examination position); Emans et al., *supra* note 36, at 784 ("Although the differences in hymenal measurements between [abused and nonabused girls] are statistically significant, the ranges overlap and . . . are unlikely to be clinically useful unless the hymen is dilated beyond the range of normal . . ."); McCann et al., *Genital Findings, supra* note 36, at 433 ("The diameters of the hymenal orifice gradually increased with the age of the child. Within each age group, the individual measurements fell along a typical bell-shaped curve . . .").

³⁸ See, e.g., Adams, *supra* note 37 (noting that normal findings include hymenal tags, hymenal bumps or mounds, anal tags in the midline, and increased perianal pigmentation); Abbey B. Berenson et al., *Appearance of the Hymen in Prepubertal Girls*, 89 PEDIATRICS 387, 394 (1992) ("Hymenal bumps, notches on the upper portion of the hymen between the 9- and 3-o'clock positions, tags, vestibular bands, longitudinal intravaginal ridges, and external ridges were all noted frequently enough in a pediatric population to be considered normal findings."); Bays & Chadwick, *supra* note 29, at 106 (referring to many of

ratio shows, symptoms that are as likely to appear in nonabused children as in abused children are not relevant for proving that abuse occurred, although they are clearly "consistent with" abuse. Expert testimony about the presence or absence of such symptoms is inappropriate and should be excluded under Federal Rule of Evidence 401.

One might object that doctors should be able to testify to the mere presence of physical symptoms "consistent with" abuse, even when the symptoms are not probative of abuse, because doctors base their diagnoses on both physical findings and the history taken from the child.³⁹ However, when a doctor states that the physical findings are "consistent with" abuse and then concludes that the child was abused, a finder of fact may erroneously conclude that the physical findings constitute independent evidence of abuse. As a result, the physician's testimony inappropriately transforms the physical evidence of uncertain relevance into corroboration for the abuse allegation.⁴⁰ In response, one might argue that such confusion can be clarified on cross-examination. This possibility exists, but the effect of a doctor's admission that the findings are also "consistent with" nonabuse may be lost on a jury, who could reasonably believe that the doctor is testifying for the prosecution because he or she believes the physical findings were caused by abuse.

Describing a child's normal condition as "consistent with" abuse transforms a doctor's belief that a child is telling the truth about

these findings as "unlikely to be due to abuse"). These research findings raise the possibility that findings made in a number of cases were not in fact attributable to abuse. See *People v. Beauchamp*, 143 A.D.2d 13, 20 (N.Y. App. Div. 1988) (Smith, J., dissenting) (The doctor found "a skin tag or extra piece of skin at the anal opening" and "stated that skin tags in the rectum without sexual abuse are rare."), *aff'd*, 539 N.E.2d 1105 (N.Y. 1989); *In re Tania J.*, 147 A.D.3d 252, 254 (N.Y. App. Div. 1989) (citing a doctor's testimony that "the anal tag and the synechia [adhesion between the hymenal ring and the labia minora] could only have resulted from chronic digital or penile penetration"); *Commonwealth v. Sullivan*, 538 A.2d 1363, 1365 (Pa. Super. Ct. 1988) (noting that a doctor "testified that he found a 'skin tag' in the child's rectal area which suggested a healed trauma consistent with penile-rectal contact"). *State v. Moore*, 749 S.W.2d 601, 602 (Tex. Ct. App. 1988) ("[The doctor] found a 'bump' at the 5 o'clock position of the hymen which is consistent with a non-accidental injury and also consistent with digital penetration of the hymen. The 'bump' is due to scar tissue forming after injury to the hymen.").

³⁹ Indeed, research documenting the existence of various physical findings among nonabused children frequently stresses that the ultimate determination whether abuse has occurred must rely in large part on the statements of the child. See McCann et al., *Genital Findings*, *supra* note 36, at 438; Bays & Chadwick, *supra* note 29, at 92. See also *Sullivan*, 538 A.2d at 1365 ("[The doctor] based his opinion that the child had been abused on a combination of social and psychological factors and said that his physical finding regarding the 'skin tag' was merely a small piece of supportive evidence.").

⁴⁰ See *In re Michelle I.*, 189 A.D.2d at 994 (holding that redness of the genitalia satisfies the requirement that hearsay be supported by "evidence tending to support the reliability" of the child's statements); *In re Alena D.*, 125 A.D.2d 753 (N.Y. App. Div. 1986) (same). As we discuss below, an irrelevant symptom does not become relevant when paired with actual evidence of abuse. See *infra* note 82 and accompanying text.

abuse into a *medical* opinion that the child was abused. Hence, what would likely be excluded as an impermissible comment on the child's credibility becomes an admissible expert opinion.⁴¹ Moreover, the testimony about an irrelevant physical finding provides the doctor an opportunity to repeat the child's statements in explaining his or her conclusion that the child was abused. Use of the "consistent with" terminology therefore enables doctors to repeat what might otherwise be excluded as inadmissible hearsay.⁴²

In sum, symptoms that are "consistent with" abuse are not necessarily relevant for proving that abuse occurred. Whereas "consistent with" commonly means that some abused children exhibit the symptom, *relevance* requires a showing that more abused children exhibit the symptom than nonabused children. Unless a proponent can demonstrate relevance, "consistent with" testimony should be excluded, not only on Rule 401 grounds, but because whatever minimal value it may have is outweighed by the possibility of unfair prejudice.

IV

ARE SYMPTOMS "CONSISTENT WITH" ABUSE RELEVANT FOR REBUTTAL?

Although testimony that a child's symptoms are "consistent with" abuse may not be relevant for proving that abuse occurred, "consistent with" testimony may nevertheless rebut the assumption that the child's condition is inconsistent with abuse. Recall that when a symptom is "consistent with" abuse, at least some abused children exhibit

⁴¹ See *State v. Wilson*, 855 P.2d 657 (Or. Ct. App. 1993). In *Wilson*, the doctor's examination revealed no physical evidence of sexual abuse. *Id.* at 658. The defendant argued that "because there was no physical evidence of abuse, the jury could infer from [Dr.] Bays' diagnosis of sexual abuse that she believed the child, making that diagnosis an impermissible comment on the child's credibility." *Id.* The court rejected the defendant's argument, noting that based on the child's interview evaluation, "*physical examination* and history, she diagnosed the child as having been sexually abused. Although, if believed, Bays' testimony supported the child's testimony, that does not render it a *direct* comment on the child's credibility. It was an opinion as to the proper medical diagnosis." *Id.* at 660 (first emphasis added). See also *State v. Butler*, 349 S.E.2d 684 (Ga. 1986). In *Butler*, the court upheld the admission of an examining physician's opinion that the child had been molested, since it "was based on her physical examination of the child as well as on the history related to her by the child." *Id.* at 685. Although the court emphasized the doctor's testimony that "the hymen was not intact and . . . the vaginal opening was larger than normal" and that "this evidence was consistent with sexual abuse," *id.*, the court of appeals noted that the doctor's "diagnosis of the child's physical condition was 'essentially normal' . . ." *Butler v. State*, 342 S.E.2d 338, 339-340 (Ga. Ct. App. 1986). But see *United States v. Whitted*, 11 F.3d 782, 786 (8th Cir. 1993) (holding that a doctor cannot express opinion that child was sexually abused when physical findings not probative of abuse).

⁴² See, e.g., *Server v. Mizell*, 902 F.2d 611, 615 (7th Cir. 1990) (holding that because of the enlarged hymenal opening in the 9-year-old, the trial court's decision to allow the child's statement to doctor, but "only as it related to the doctor's diagnosis of sexual assault, was proper").

the symptom. In contrast, if a symptom is inconsistent with abuse, no abused children exhibit the symptom. Therefore, testimony that a child's symptom is "consistent with" abuse rebuts the assumption that no abused children exhibit the symptom. In terms of the relevance ratio, testimony that a child's symptom is "consistent with" abuse rebuts the assumption that the numerator of the ratio is zero, but says nothing about the denominator.

Consider a case in which the defense introduces evidence suggesting that abuse did not occur. For example, the defense may refer to a substantial delay in time between the alleged abuse and the child's first report, or the defense may stress the failure of the child's physical examination to reveal any physical signs of abuse. If the jurors are inclined to assume that all abused children report abuse promptly or that all abused children show physical signs of abuse, the facts emphasized by the defense appear inconsistent with the allegation that abuse occurred. Testimony that some abused children delay in reporting the abuse⁴³ and that some abused children show no physical signs of abuse⁴⁴ rebuts this assumption. Even if the jury understands that not all abused children promptly report abuse or show physical signs of abuse, the jury may still overestimate the percentage of abused children who do so. Thus, testimony about delayed reporting of abuse or that a lack of physical signs is common among abused children can help to overcome the false presumption that such evidence is virtually non-existent among abused children.

Most courts allow prosecution experts to testify that an alleged victim's behavior is "consistent with" abuse to rebut defense claims that the behavior proves abuse did not occur.⁴⁵ Although this reasoning is generally sound, courts must emphasize the limited purpose of such testimony. Rebuttal testimony that a symptom is "consistent with" abuse is not the same as testimony that a symptom is relevant for proving that abuse occurred. In terms of the relevance ratio, rebuttal testimony is equivalent to a claim that the numerator of the relevance ratio is greater than zero, whereas testimony offered to prove that abuse occurred is equivalent to a claim that the ratio is greater than one.

⁴³ See, e.g., *People v. Gilbert*, 7 Cal. Rptr. 2d 660, 667 (Cal. Ct. App. 1992).

⁴⁴ See, e.g., *People v. Rowland*, 841 P.2d 897, 914 (Cal. 1992) (allowing doctor's testimony that "the absence of genital trauma is not inconsistent with nonconsensual sexual intercourse"); Susan Morison & Edith Greene, *Juror and Expert Knowledge of Child Sexual Abuse*, 16 CHILD ABUSE & NEGLECT 595, 608 (1992) (finding that over half of the lay survey respondents believed that "most children are physically damaged" by sexual abuse).

⁴⁵ See JOHN E.B. MYERS, EVIDENCE IN CHILD ABUSE AND NEGLECT CASES 292 (2d ed. 1992) (discussing the use of rape trauma syndrome evidence as rebuttal testimony); *id.* at 310 (discussing rebuttal testimony about the behavior of child sexual abuse victims).

Nevertheless, some courts and commentators have not appreciated the difference between these two claims. For example, the New Mexico Supreme Court reasoned that if admitting rebuttal testimony that a child's behavior is "consistent with" abuse is permissible, admitting testimony that a child's behavior is "consistent with" abuse to prove that abuse occurred should also be permissible.⁴⁶ Similarly fallacious inferences appear in the literature among those who start with the opposite premise. For example, commentator Robert Levy argues that because symptoms of abuse are not relevant for proving that abuse occurred, symptoms cannot be relevant in rebuttal.⁴⁷ In both cases, the error results from conflating that which is consistent with abuse with that which is relevant for proving that abuse occurred.

Another error occurs when rebuttal testimony that a child's symptoms are "consistent with" abuse is treated as identical to testimony that those symptoms are irrelevant to determining whether abuse occurred. As noted above, rebuttal testimony merely demonstrates that the numerator of the relevance ratio is nonzero (i.e., that at least some abused children exhibit the symptom in question). But, even with a numerator greater than zero, the evidence may be relevant evidence that a child was not abused, if a larger proportion of nonabused children exhibit the symptom than abused children. Thus, paradoxically, the same item of evidence may be valuable to the prosecution for rebuttal purposes *and* to the defense as proof of nonabuse.

For example, a court may properly admit testimony that many abused children have normal genitalia to rebut the assumption that abused children invariably show physical signs of abuse. Testimony of this sort demonstrates that the number of abused children with normal genitalia is greater than zero. However, it is likely that the pro-

⁴⁶ See *State v. Alberico*, 861 P.2d 192, 210 (N.M. 1993) ("Allowing an expert to testify that PTSD [post traumatic stress disorder] symptoms are a common reaction to sexual assault for the purpose of rebutting the defense that the victim's reaction to the alleged incident are inconsistent with sexual assault is no different from allowing the expert to testify that the alleged victim's symptoms are consistent with sexual abuse."). Note that the court refers to testimony that certain symptoms are a "common" reaction to abuse; the significance of symptoms that are "common" rather than merely "consistent" is discussed *infra* text accompanying notes 51-66. See also *Steward v. State*, 652 N.E.2d. 490, 501 (Ind. 1995) (Sullivan, J., dissenting) ("[W]e can[not] say that the scientific principles upon which child sexual abuse accommodation syndrome is based are not reliable as direct evidence but are reliable as rebuttal evidence. In logic, either those principles are reliable for . . . aiding the jury in assessing the child victim's credibility or they are not").

⁴⁷ See Levy, *supra* note 6, at 394-95 (criticizing the defenders of the rehabilitative use of child sexual abuse accommodation syndrome for lacking "an explanation of how a set of symptoms that have no empirical justification for proving abuse can nonetheless be used to prove that the child's recantation is false"). Cf. Cohen, *supra* note 6, at 444, 446. Cohen objects to testimony that retracting allegations is a typical behavior for sexual abuse victims because "[t]here is something fundamentally strange about saying that since the child denies that the event occurred, it must have occurred." The error is in assuming that rebuttal testimony suggests that abuse "must have" occurred, rather than "may have" occurred.

portion of abused children with normal genitalia, albeit greater than zero, is less than the proportion of nonabused children with normal genitalia.⁴⁸ Consequently, normal genitalia are "consistent with" abuse, but, by a relevance ratio analysis, are also evidence that abuse did not occur.⁴⁹ The apparently paradoxical nature of this conclusion disappears when one recognizes that evidence that is consistent with one side's theory of a case may also be consistent with the other side's theory. A relevance ratio analysis indicates which theory receives greater support from such evidence. In this example, the evidence should disabuse the jurors of the presumption that a lack of physical signs conclusively disproves abuse, but the evidence should not lead juries to believe that the lack of physical signs is totally irrelevant.⁵⁰

The distinction between testimony that tends to prove that abuse occurred and testimony that rebuts the assertion that abuse did not occur clarifies some of the confusion about limitations on expert testimony in child abuse cases. Some courts allow rebuttal testimony, but exclude proof-of-abuse testimony on the ground that proof-of-abuse testimony constitutes an impermissible comment on the credibility of the child.⁵¹ This rationale leaves something to be desired. Both types of testimony presuppose that the jury needs to be informed regarding the probativeness of symptoms of abuse. Proof-of-abuse testimony informs the jury that a symptom is more common among abused children than among nonabused children, whereas rebuttal testimony informs the jury that a symptom is more common among abused children than the jury might otherwise have supposed (and, therefore, less probative of nonabuse than the jury might believe). Therefore, to the extent rebuttal testimony "invades the province of the jury" by telling jurors how to weigh the child's credibility, rebuttal testimony is as objectionable as proof-of-abuse testimony.⁵² In terms of the relevance ratio, however, evidence that is relevant for rebuttal is not necessarily relevant for proving that abuse occurred. In other words, the ratio allows courts to make a sensible distinction between the two ways in which the evidence may be used.

⁴⁸ This follows from the assumption that the proportion of abused children with abnormal genitalia is greater than the proportion of nonabused children with abnormal genitalia.

⁴⁹ Rebuttal experts will frequently note that lacking signs of abuse is not only consistent with abuse, but common among abused children. Nevertheless, lacking a sign may still be evidence that abuse did not occur, if nonabused children are more likely to lack signs than abused children.

⁵⁰ In this example, the relevance ratio is greater than zero because the ratio includes a non-zero numerator, but is less than one because the denominator is greater than the numerator.

⁵¹ See, e.g., *State v. Moran*, 728 P.2d 248, 254-55 (Ariz. 1986).

⁵² Indeed, this is the conclusion reached by the Pennsylvania Supreme Court. See *Commonwealth v. Dunkle*, 602 A.2d 830, 837 (Pa. 1992).

In sum, testimony that a symptom is "consistent with" abuse is relevant evidence when offered to rebut the misconception that the symptom is nonexistent or uncommon among abused children. Such testimony may disabuse a jury of the assumption that a symptom is inconsistent with abuse and, therefore, conclusively disproves the abuse hypothesis. However, the court must take care to ensure that a jury is not misled into believing that the symptom is either irrelevant or tends to prove that abuse *did* occur. Even if the symptom such as normal genitalia occurs with some frequency among abused children, it may still tend to prove that abuse did *not* occur if the defendant can show that the symptom is even more common among nonabused children.

V

ARE COMMON SYMPTOMS RELEVANT FOR PROVING ABUSE?

A psychologist's first priority when treating a sexually abused child is the child's well-being. Effective treatment requires not only an ability to analyze and understand the unique features of each child's case, but also a thorough knowledge of the emotional and behavioral patterns of abused children. Once a therapist understands these patterns, the therapist is in a better position to anticipate problems and to prepare the child for the rough road that may lie ahead. As research on common symptoms and patterns progresses, the treatment of abused children will continue to improve.

Whereas identification of symptoms common among abused children has therapeutic value, it may or may not have value for the question of whether abuse occurred. What is common among abused children and what is relevant for proving abuse are different. Common symptoms of abuse are relevant for proving abuse only if they are less commonly observed in nonabused children. By the same token, even uncommon symptoms of abuse may be highly relevant, if they are still more uncommon among nonabused children. In short, the frequency of a symptom is not synonymous with probativeness or relevance. It is their relative frequency, as measured by the relevance ratio, that indicates whether and to what extent a symptom is relevant for proving that abuse occurred.

Research on the behavioral symptoms of sexually abused children illustrates that common symptoms may be irrelevant and uncommon symptoms may be relevant for proving abuse. Hibbard and Hartman compared sexually abused children to nonabused children using a number of behavioral and emotional variables.⁵³ The authors found

⁵³ Roberta A. Hibbard & Georgia L. Hartman, *Behavioral Problems in Alleged Sexual Abuse Victims*, 16 CHILD ABUSE & NEGLECT 755 (1992). Their study included 171 4-8 year old children, 81 of whom were allegedly sexually abused, and 90 of whom were receiving

that nightmares were one of the most common complaints, whereas sexual problems and sex play were relatively less common. However, they also found that nightmares are quite common among nonabused children (in fact, they are slightly, but not significantly, *more* common among nonabused children), whereas sexual problems and sex play are virtually nonexistent in this group. Although an emphasis on frequency would lead one to focus on symptoms that have little probative value, a relevance ratio analysis reveals that sexual behavior may be one of the most probative symptoms of sexual abuse. Subsequent research supports the conclusion that certain types of highly sexualized behavior (e.g., solicitations to engage in sexual acts and masturbation with an object) are uncommon among abused children, but are even more uncommon among children who have not been abused.⁵⁴ Thus, despite being uncommon among abused children, sexual behavior is quite relevant for proving that abuse occurred.

The medical literature on sexual abuse contains similar examples of common symptoms appearing irrelevant and uncommon symptoms appearing relevant. The most common physical symptom of sexual abuse is redness of the genitalia (erythema).⁵⁵ Doctors frequently mention erythema in their testimony about abuse.⁵⁶ However, because erythema does not appear to be more common among abused children than among nonabused children, it is not relevant for proving that abuse occurred and should not be offered for this purpose.⁵⁷ In contrast, doctors observe gonorrhea in less than 5% of sexually abused children.⁵⁸ But, because gonorrhea is virtually non-existent among nonabused children who have not had sexual contact, it is strong evidence that abuse has occurred.⁵⁹

well-child care in a pediatric clinic; each child's parents completed the Child Behavior Checklist; 41% of abused and 42% of nonabused reported nightmares; 19% of abused and 1% of nonabused reported sexual problems; 19% of abused and 0% of nonabused reported too much sex play; 11% of abused and 3% of nonabused reported sex play in public. *Id.* at 758. One should interpret these figures with caution, however, given the potential for "detection biases" and other methodological problems, as discussed *infra* text accompanying notes 86-96.

⁵⁴ Friedrich et al. found a number of sexual behaviors that occurred among about 10% of abused children, but less than 1% of nonabused children, in a sample of 276 abused children and 880 nonabused children ages to twelve years. The behaviors included "ask[ing] to engage in sex acts," "masturbat[ing] with object[s]," and "insert[ing] objects in vagina/anus." William N. Friedrich et al., *Child Sexual Behavior Inventory: Normative and Clinical Comparisons*, 4 PSYCHOL. ASSESSMENT 303, 307 (1992).

⁵⁵ See, e.g., Emans et al., *supra* note 36, at 782 (concluding that erythema is the most common physical finding among sexually abused sample).

⁵⁶ See *supra* notes 29-32.

⁵⁷ See *supra* note 36.

⁵⁸ See Bays & Chadwick, *supra* note 29, at 93 (reviewing six studies that found sexually transmitted diseases in 2% to 7% of abused children).

⁵⁹ See Deborah Stewart, *Sexually Transmitted Diseases*, in EVALUATION OF THE SEXUALLY ABUSED CHILD: A MEDICAL TEXTBOOK AND PHOTOGRAPHIC ATLAS 145, 150 (1992). If per-

These examples highlight the difference between symptoms that are common among abused children and symptoms that are relevant for proving that a child has been abused. Symptoms found among abused children may be irrelevant for diagnosing abuse if those symptoms also occur among a broad range of nonabused children.⁶⁰

A misguided focus on the typicality of symptoms has caused courts to err in two distinct directions. Some courts relying on the commonality of the symptom have allowed expert testimony regarding symptoms of abuse as a means of proving that abuse occurred.⁶¹ Other courts and some commentators have opposed admitting expert

inatal transmission is ruled out, gonorrhea is considered diagnostic of sexual abuse. See Adams, *supra* note 37, at 13; American Academy of Pediatrics, Committee on Child Abuse and Neglect, *Guidelines for the Evaluation of Sexual Abuse of Children*, 87 PEDIATRICS 254, 257 (1991); Bays & Chadwick, *supra* note 29.

⁶⁰ This is not to say that symptoms used for the purposes of treating sexual abuse are necessarily irrelevant in proving sexual abuse, as some courts and commentators have concluded. See, e.g., *People v. Beckley*, 456 N.W.2d 391, 405 (Mich. 1990) ("Child sexual abuse syndrome is not intended as a diagnostic tool for detection of sexual abuse. Thus, it has no probative value in terms of being able to detect sexual abuse on the basis of the existence of certain behavioral characteristics."); Askowitz & Graham, *supra* note 6, at 2046 ("Post-traumatic stress disorder merely is a therapeutic tool; it is not designed to determine sexual abuse. In practice, however, experts ignore these limitations of the PTSD diagnosis and use it to opine that the child has been sexually abused."). For example, children who have been sexually abused are often tested for gonorrhea, so that if they have been infected, they may be treated. That the presence of gonorrhea is useful in determining treatment says nothing about the probative value of gonorrhea. Coincidentally, however, gonorrhea is highly probative of abuse. See Lawrence S. Neinstein et al., *Nonsexual Transmission of Sexually Transmitted Diseases: An Infrequent Occurrence*, 74 PEDIATRICS 67, 67-68 (1984).

⁶¹ See *United States v. Snipes*, 18 M.J. 172, 179 (1984) (allowing an expert to conclude that the child was abused, because there is a "sufficient body of 'specialized knowledge' as to the typical behavior of sexually abused children"); *Ward v. State*, 519 So. 2d 1082, 1083 (Fla. Dist. Ct. App. 1988) (allowing expert testimony that the child "displayed the symptoms typically seen in children who have been sexually abused"); *State v. Reser*, 767 P.2d 1277, 1280 (Kan. 1989) (allowing expert testimony that the child "exhibits some of the characteristics commonly found in sexually abused children"); *State v. Donnelly*, 798 P.2d 89, 91 (Mont. 1990) (allowing expert testimony "concerning general symptoms often found in children who have been sexually abused"), *rev'd on other grounds*, 813 P.2d 979 (Mont. 1991); *State v. Schumpert*, 435 S.E.2d 859, 861 (S.C. 1993) (approving expert testimony that "the victim's behavioral symptoms were typical for a victim of sexual abuse"); *State v. Stevens*, 794 P.2d 38, 41 (Wash. Ct. App. 1990) (allowing expert testimony "that sexually abused children exhibit common behaviors"); *Rivera v. State*, 840 P.2d 933, 939 (Wyo. 1992) (approving expert testimony "explaining to the jury the typical behavioral patterns of adolescent victims of sexual assault"). Testimony as to the typical characteristics of child abuse victims has been admitted by other courts without explicit reference to the "typicality" of symptoms. See *State v. Hammond*, 435 S.E.2d 798, 802 (N.C. Ct. App. 1993) ("It was proper for [the expert] to discuss the symptoms and characteristics of sexually abused children and to express, in her expert opinion, whether the minor child exhibited similar characteristics."); *State v. Edward Charles L.*, 398 S.E.2d 123, 139 (W.Va. 1990) (allowing expert testimony that a child had been abused, because "children who are sexually abused or assaulted frequently display indications which may comport to a profile of child victims").

testimony on the consequences of abuse arguing that such symptoms are so uncommon that they are of limited relevance.⁶²

For example, in *State v. Foret*, the Supreme Court of Louisiana held that a finding that 68% of abused children suffer from certain symptoms of abuse is not sufficient to render these symptoms relevant and probative.⁶³ The court reasoned that if 32% of abused children do not exhibit these symptoms (a proportion that the court identified as a "32% margin of error"), testimony about the presence of these symptoms is "unreliable."⁶⁴ But, for reasons articulated throughout this Article, this logic does not withstand careful scrutiny. An assessment of probative value does not require a comparison of the proportion of abused who exhibit the symptoms to the proportion of *abused* children who do not exhibit the symptom. Nor does a conclusion that the symptom is probative of abuse require that the proportion of abused who do not exhibit the symptom be less than some predetermined "margin of error." Instead, the probative value is determined by the ratio of the properties of abused children who exhibit the symptom to the proportion of *nonabused* children who exhibit the symptom. If the symptom occurs more frequently among abused children than among nonabused children, yielding a relevance ratio greater than one, the symptom is relevant for proving that abuse occurred.

The Pennsylvania Supreme Court also surrendered to the temptation to use a symptom's typicality in the abused population as a proxy for assessing its relevance and probative value. In *Commonwealth v. Dunkle*, the court concluded that because "abused children cannot be fit into any specific behavior patterns . . . [and because] not one, single symptom was exhibited by a majority of sexually abused children . . . [c]learly, these types of percentages cannot constitute probative evidence."⁶⁵ But probative value does not require that the symptom in question occur among a majority or even a large proportion of abused

⁶² See Askowitz & Graham, *supra* note 6, at 2085-86 (arguing that there are no typical symptoms of child abuse); Lisa R. Askowitz, Comment, *Restricting the Admissibility of Expert Testimony in Child Sexual Abuse Prosecutions: Pennsylvania Takes it to the Extreme*, 47 U. MIAMI L. REV. 201, 208-09 (1992) ("[S]tudies have revealed the absence of a single, typical child abuse victim In fact, almost every study of the impact of sexual abuse on children has found a substantial group of victims with little or no symptomatology."); Cohen, *supra* note 6, at 440 (arguing that there are no typical consequences of abuse); McCord, *supra* note 6, at 19-20 ("Because of the variability of responses, researchers have been frank to admit that they simply cannot tell what the reactions to sexual abuse of any particular child will be. . . . Since there are no typical symptoms, diagnosis is very difficult.").

⁶³ 628 So. 2d 1116, 1126 (La. 1993).

⁶⁴ *Id.*

⁶⁵ 602 A.2d 830, 835 n.16 (Pa. 1992). See also *Gier v. Educational Serv. Unit No. 16*, 845 F. Supp. 1342, 1348 n.10 (D. Neb. 1994) (quoting *Dunkle*, 602 A.2d at 835 n.16), *aff'd*, 66 F.3d 940 (8th Cir. 1995). For a discussion of *Dunkle's* treatment of the denominator of the relevance ratio, see *infra* text accompanying notes 108-09.

children.⁶⁶ Nor does probative value require that the symptom be absent from nonabused children. As indicated throughout this Article, all that is necessary is a greater frequency of the symptom among the abused than among the nonabused population.

Dunkle is also noteworthy for its discussion of the court's understanding of relevance. This elaboration contains some serious shortcomings. First, the court notes that relevant evidence is evidence that makes a fact in issue more probable than it would be without the evidence.⁶⁷ This much is correct. But, the court subsequently paraphrases this standard as a requirement that relevant evidence make a fact in issue "more probable than not."⁶⁸ Here the court actually redefines, rather than paraphrases, the relevance standard. A symptom may make abuse more probable without making it "more probable than not." A symptom that doubles the probability of abuse from 20% to 40% is relevant under a relevance ratio analysis; the symptom, however, would not be relevant by the "more probable than not" standard imposed by the Pennsylvania Court. Commentators have discussed the confusion between evidence that makes a fact in issue "more probable" versus "more probable than not" in other contexts as well.⁶⁹

The Pennsylvania Court's insistence that a symptom can make abuse "more likely than not" only when it is present in more than half of all abused children exacerbates its confusion. This requirement is neither necessary nor sufficient for a symptom to make abuse "more likely than not." If a symptom appears among few abused children, but among still fewer nonabused children, that symptom could make abuse "more likely than not"; conversely, if a symptom appeared among most abused children, but among somewhat fewer nonabused children, that symptom may nonetheless fail to make abuse "more likely than not."

Both proponents and critics of behavioral science testimony in child sexual abuse cases have fallen prey to the error of conflating commonality with relevance. Familiarity with the relevance ratio would enable all parties, including trial judges who must make admissibility decisions, to identify the degree of probative value of a wide

⁶⁶ See *supra* text accompanying notes 58-59 (presence of gonorrhea is probative of abuse, because it appears in five percent of abused children, but in an even smaller proportion of nonabused children).

⁶⁷ *Dunkle*, 602 A.2d at 834.

⁶⁸ *Id.*

⁶⁹ See JOHN W. STRONG ET AL., MCCORMICK ON EVIDENCE § 185, at 776-77 (4th ed. 1992); see also *People v. Green*, 573 N.Y.S.2d 113, 116 (N.Y. Sup. Ct. 1991) ("[a]ssuming *arguendo*, that defendant is correct that the probability he was accurately identified is less than 1/2, it does not necessarily follow that the identification testimony could not, with proper instructions, be submitted to the jury [T]he thrust of defendant's argument affects the *weight* of identification testimony and not its admissibility.").

range of symptoms. Many common symptoms are only marginally relevant, and many uncommon symptoms have great probative value.

VI

ARE CLUSTERS OF SYMPTOMS RELEVANT FOR PROVING ABUSE?

Some argue that even though individual symptoms of abuse may not be probative, the appearance of multiple symptoms (i.e., a cluster of symptoms) is highly suggestive of abuse.⁷⁰ Presumably, the appearance of a cluster of symptoms is uncommon among nonabused children. Although this position has intuitive appeal, clusters of symptoms may be far less probative than they at first appear to be. There are several reasons why this is true.

First, a cluster of symptoms is almost certain to be more common than a single symptom among *both* abused and nonabused populations. This increased frequency occurs largely because of the vast number of symptoms that are associated with abuse. Assume, for example, that there are ten symptoms of abuse.⁷¹ There are only ten ways in which a single symptom alone might present itself (e.g., symptom #1 only, symptom #2 only, . . . , symptom #10 only). However, there are more than one thousand ways in which two or more symptoms may appear (e.g., symptoms #1 and #2 only, symptoms #1, #2, and #3 only, etc.).⁷² Unless all of the individual symptoms are exceedingly rare, the probability that a cluster will present itself in a nonabused child will be substantially greater than the probability that a single symptom will present itself. In other words, because so many

⁷⁰ See, e.g., *State v. Michaels*, 625 A.2d 489, 497 (N.J. Super. Ct. App. Div. 1993) (noting that the expert testified "that what is crucial is the clustering of those [behavioral] symptoms" along with other indicators of abuse), *aff'd*, 642 A.2d 1372; *State v. Hammond*, 435 S.E.2d 798, 802 (N.C. Ct. App. 1993) (explaining that the expert testified that "none of these symptoms by itself is a total indication that sexual abuse is present, but I think with the clustering of these symptoms that [the minor child] showed, that there is a very high probability that she had been sexually abused.").

⁷¹ This is a conservative estimate. See *Michaels*, 625 A.2d at 496 (noting that the expert testified that she used a list of 32 behaviors.); *Dunkle*, 602 A.2d at 835 n.16 (reviewing study that examined 38 behaviors); Friedrich et al., *supra* note 54, at 304 (noting that the Child Sexual Behavior Inventory assesses 35 sexualized behaviors); Kendall-Tackett et al., *supra* note 19, at 166 (reviewing research on 24 symptoms).

⁷² The number of clusters of symptoms (i.e., sets of two or more symptoms) associated with a set of n symptoms is given by the following formula:

$$2^n - \binom{n}{1} - \binom{n}{0}.$$

Thus, there are

$$2^{10} - \binom{10}{1} - \binom{10}{0} = 1,013$$

clusters possible from a set of 10 possible symptoms.

different possibilities exist, one should not be surprised to find clusters of symptoms among children who have not been abused.

To make this argument more concrete, consider the following example. Assume that each of the ten symptoms appears among twenty percent of nonabused children. Assume further that the symptoms are conditionally independent. This means that the presence or absence of any one of the ten symptoms does not affect the likelihood that any of the remaining nine symptoms will present themselves.⁷³ The probability that a nonabused child will exhibit one symptom is only 27%.⁷⁴ However, the probability that a child will exhibit two or more symptoms is a hefty 62%.⁷⁵ Of course, the probability associated with the presence of exactly one symptom or the presence of multiple symptoms is a function both of the individual symptom's probability and the number of possible symptoms. But, under most reasonable sets of assumptions about these values, the probability that a nonabused child will exhibit a cluster of symptoms is substantially greater than the probability that this child will exhibit only a single symptom.⁷⁶ That clusters of symptoms may be common among children who have not been abused has important implications for an assessment of the probative value of symptom clusters. Even if an extremely large proportion of abused children exhibit a particular cluster of symptoms, this cluster may be of minimal diagnostic value if a large proportion of nonabused children also exhibit this cluster of symptoms. Thus, the intuition that the presence of multiple symptoms must be diagnostic of abuse vanishes under the scrutiny of relevance ratio analysis.

We suspect that the false belief that clusters of symptoms must be probative of abuse is in part attributable to a failure to distinguish between the probability that a *particular* cluster of symptoms will occur, a rare event, and the probability that a cluster of symptoms of *some sort* will occur (a frequent event).⁷⁷ Returning to our example, suppose a child displays symptoms 1, 2, and 3, but does not display

⁷³ More formally, symptoms A and B are conditionally independent if $P(A)P(B) = P(A|B)P(B)$.

⁷⁴

$$P(\text{Single symptom}) = \binom{10}{1} (.2)^1 (.8)^9 = .27$$

⁷⁵ $P(2 \text{ or more symptoms}) = 1 - \binom{10}{1} (.2)^1 (.8)^9 - \binom{10}{0} (.2)^0 (.8)^{10} = 1 - .38 = .62$.

$$P(\text{Single symptom}) = 1 - \binom{10}{0} (.2)^0 (.8)^{10} - \binom{10}{1} (.2)^1 (.8)^9 = .62$$

⁷⁶ Exceptions to this generalization occur when few possible symptoms exist or when those symptoms are rare.

⁷⁷ A similar phenomenon underlies people's intuitions in the famous "birthday problem." See THOMAS GILOVICH, *HOW WE KNOW WHAT ISN'T SO* (1991). Gilovich writes: When asked to consider the probability that at least two people in a group of a particular size were born on the same day of the year, most people are

symptoms 4 through 10. Although the probability that a nonabused child would exhibit this particular cluster of three symptoms is rare, less than 0.2%,⁷⁸ the probability that a nonabused child would exhibit *some* cluster of three symptoms is a surprisingly high 20.1%.⁷⁹ This dramatic increase is attributable to the large number of ways in which clusters of three or more symptoms can occur when the list of possible symptoms is large (i.e., as great as ten). Unless an expert has specified in advance the particular clusters he or she considers most significant, *any* cluster that presents itself, in retrospect, will likely appear suggestive of abuse.

Belief in the probative value of clusters may also persist because of the correlated nature of the symptoms. The ten-symptoms hypothetical assumed that each symptom appeared independently of the other nine symptoms. This assumption simplified the computations and illustrated the point that clusters of symptoms may be common among nonabused children. However, it is likely that many abuse symptoms tend to occur contemporaneously, among both abused and nonabused children.⁸⁰ To the extent this is true, these symptoms are likely to cluster among both abused and nonabused children. In such a case, the presence of multiple symptoms in a cluster may tell one little more than if a single symptom existed. The additional symptoms are essentially redundant information and do not help to distinguish between abused and nonabused children.⁸¹

shocked to learn that the odds are roughly 50-50 when the group is as small as 23. . . .

. . . .

. . . Many people approach the problem with a fairly accurate sense of the long odds against a *particular* pair of people having the same birthday (approximately 1/365), but they fail to appreciate how many different pairs of people there are (253) in a group of 23.

Id. at 175-77.

⁷⁸ $P(1-2-3 \text{ cluster}) = (.2)^3 (.8)^7 = .0017$.

⁷⁹ $P(\text{cluster of 3 symptoms}) =$

$$\binom{10}{3} (.2)^3 (.8)^7 = .201$$

Other things being equal, the appearance of clusters of symptoms will become more common as the number of possible symptoms increases.

⁸⁰ Symptoms examined in the research on behavioral reactions to abuse are often very similar (e.g., anxious/feels too guilty/worries or trouble sleeping/nightmares). See, e.g., Hibbard & Hartman, *supra* note 53, at 758. Any child who suffers one of these related symptoms is more likely to suffer other similar symptoms, whether or not that child has been abused.

⁸¹ Research suggests that people do not understand that confidence in one's judgments should decrease, rather than increase, when the predictors are highly correlated. Kahneman and Tversky called the tendency to become increasingly confident in the face of correlated predictors the "illusion of validity." Daniel Kahneman & Amos Tversky, *On the Psychology of Prediction*, in *JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES* 48, 65-66 (Daniel Kahneman et al. eds., 1982).

Some experts have argued that "allegations of sexual abuse combined with certain other behavioral reactions make the 'non-probative' reactions more probative of sexual abuse in the clinical context."⁸² This argument does not withstand scrutiny. Not only are there no data to support this empirical claim, but it is hard to imagine how a symptom that is not relevant for proving that abuse occurred becomes relevant simply by being paired with a symptom that is relevant.

On the other hand, many situations exist in which clusters of symptoms *are* substantially more probative than single symptoms. In abuse cases involving infants, skeletal fractures at different stages of healing constitute highly probative evidence of abuse, whereas a single skeletal fracture, standing alone, provides only moderately probative evidence of abuse.⁸³ Children with fractures who were abused are more likely than accidentally injured children to have been seriously injured at a previous time, making multiple fractures much more common among abused than nonabused children.⁸⁴

In sum, intuition may not provide a reliable method for determining the probative value of symptom clusters.⁸⁵ The observation that a particular symptom cluster (e.g., nightmares + anxiety + low self-esteem) occurs more frequently in abused populations than any of the individual symptoms alone does not demonstrate relevance. To demonstrate relevance, research must show that the particular cluster is more likely to appear in abused children than in nonabused children. To date, remarkably little data exist that support the strong

⁸² Patricia A. Korey, *Evidence—Rehabilitative Expert Testimony in Child Sexual Abuse Cases: The Supreme Court of Pennsylvania Shuts the Door on Effective Prosecutions—Commonwealth v. Dunkle*, 602 A.2d 830 (Pa. 1992), 66 TEMPLE L. REV. 589, 605 (1992). Korey's statement is understandable, given the premise that what makes some symptoms non-probative is that they may be attributable to causes other than sexual abuse. *Id.* at 605-06. Korey's statement, however, is itself untrue. The possibility of alternate causes indicates that some nonabused children will exhibit the symptoms, but does not mean that the symptoms are therefore irrelevant. The article on which Korey bases her argument clarifies this point. The paper concludes that "[w]hen symptoms of fear, anxiety, or avoidance accompany a credible report of sexual abuse, sexual abuse must be seriously considered." John E.B. Myers et al., *Expert Testimony in Child Sexual Abuse Litigation*, 68 NEB. L. REV. 1, 61 (1989). Myers et al. do not in fact argue that such symptoms are non-probative of abuse; rather, they comment upon the possibility that the symptoms are attributable to other traumatic events besides abuse. *Id.* at 60-61. Subsequently, they clarify that an expert should not "describe a combination of behaviors, some of which are probative of abuse and some of which are not." *Id.* at 69.

⁸³ See Helen Carty, *Brittle or Battered*, 63 ARCHIVES OF DISEASE IN CHILDREN 350, 350 (1988); C.J. Hobbs, *Fractures*, 298 BRIT. J. MED. 1015, 1015 (1989).

⁸⁴ See Peter Worlock et al., *Patterns of Fractures in Accidental and Non-Accidental Injury in Children: A Comparative Study*, 293 BRIT. MED. J. 100, 101 (1986) (studying 35 abused children and 116 non-abused children with fractures and finding abused children much more likely than nonabused children to have multiple fractures).

⁸⁵ Cf. Koehler & Shaviro, *supra* note 23, at 266-72.

claims concerning the probative value of clusters made by some clinicians at trial.

VII

ARE ALL STUDIES CREATED EQUAL?

Although assessing the relevance of possible symptoms of abuse requires studies that compare abused and nonabused children, not all such studies are equally informative. Methodological shortcomings exist in some studies that falsely inflate the probative value of certain symptoms of abuse.

The most serious methodological shortcoming that courts should guard against is "selection bias."⁸⁶ In the context of abuse studies, selection bias occurs when the abused and nonabused children who are selected for study differ in ways that may affect the variables measured in the studies. For example, suppose a study draws a sample of abused children from among those who are in therapy for psychological problems, but draws a nonabused comparison sample from among children who are not in therapy. In such a study, a substantially higher proportion of the abused children will likely exhibit psychological problems, even problems that are not related to abuse.⁸⁷

One way to redress this problem is to compare abused children who are in therapy with nonabused children who are in therapy for problems other than abuse.⁸⁸ In this type of therapy-matched study, one can be more confident that the problems that are more frequent among the abused children than the nonabused children are diagnostic of abuse.⁸⁹ A review of these studies found that many symptoms

⁸⁶ See HARVEY CHECKOWAY ET AL., RESEARCH METHODS IN OCCUPATIONAL EPIDEMIOLOGY 77 (1989) ("Selection bias is any bias arising from the procedures by which the study subjects were chosen from the entire population that theoretically could be studied.").

⁸⁷ See Kendall-Tackett et al., *supra* note 19, at 165 (stating that most of studies reviewed on the effects of sexual abuse "primarily drew from sexual abuse evaluation or treatment programs"). The researchers themselves frequently recognize the problem this presents. See, e.g., Alison J. Einbender & William N. Friedrich, *Psychological Functioning and Behavior of Sexually Abused Girls*, 57 J. CONSULTING & CLINICAL PSYCHOL. 155, 155 (1989) ("[D]ifferences between these two groups of girls [abused and nonabused] may have existed as a result of selection criteria for the sexually abused girls, most of whom were referred to the study by clinicians."); William N. Friedrich et al., *Behavior Problems in Sexually Abused Young Children*, 11 J. PEDIATRIC PSYCHOL. 47, 49 (1986) ("Due to the fact that children demonstrating more problems are more likely to be referred for treatment, as a group these children are probably displaying a more pronounced response to the abuse.").

⁸⁸ Another solution is to draw both the abused and nonabused samples from the population of children in general, rather than from among those in therapy. Unfortunately, this solution is difficult to implement, because it is hard to identify and recruit abused children who are not in treatment.

⁸⁹ To the extent that children are referred to treatment solely because they were sexually abused, and not because of psychological problems, these comparisons may underestimate the effects of sexual abuse on psychological functioning. See Kendall-Tackett et al., *supra* note 19, at 165. They state:

(e.g., aggressive behavior, hyperactivity) do not occur more commonly among abused children in therapy than among nonabused children in therapy.⁹⁰ On the other hand, some symptoms appear more frequently among abused children, such as sexualized behavior.⁹¹

"Detection bias" is a second methodological problem. "Detection bias" occurs when the presence of a symptom increases the chances that abuse will be detected.⁹² Although observing symptoms is a useful strategy for detecting abuse, one must then exercise care when assessing the diagnostic strength of these symptoms. If children are classified as abused or nonabused on the basis of certain symptoms, it is circular to estimate the diagnostic value of those symptoms based on children thus classified.⁹³ If one uses physical signs as a means of dis-

[C]linical comparison groups generally contain many children who are referred primarily because of their symptomatic behavior. Naturally these children are likely to be more symptomatic than children referred not because of symptoms, but because of something done to them (i.e., abuse). Thus, the lower levels of symptoms in sexually abused children may say more about the clinical comparisons than about the sexually abused children themselves.

Id.

⁹⁰ See Kendall-Tackett et al., *supra* note 19, at 165-66.

⁹¹ In a review of the research Kendall-Tackett et al. found that sexually abused children consistently exhibited sexualized behavior more often than nonabused children. Kendall-Tackett et al., *supra* note 19, at 165-66. See also Esther Deblinger et al., *Post-Traumatic Stress in Sexually Abused, Physically Abused, and Nonabused Children*, 13 CHILD ABUSE & NEGLECT 403, 406 (1989) (finding that in a retrospective chart review of medical inpatients at psychiatric facility (29 sexually abused, 29 physically abused, and 29 nonabused children), "[s]exually abused children . . . suffered significantly more reexperiencing phenomena than either physically abused or nonabused subjects [and that an] examination of the symptoms within this PTSD subcategory revealed that this was attributable to the significantly higher rate of inappropriate sexual behaviors displayed by child sexual abuse victims."); William N. Friedrich, *The Clinical Use of the Child Sexual Behavior Inventory: Commonly Asked Questions*, 8 APSAC ADVISOR 1, 18 (1995) (stating that a preliminary analysis of 100 psychiatric patients found that although they had elevated levels of sexualized behavior relative to normal children, they had lower levels than group of sexually abused children.); John Gale et al., *Sexual Abuse in Young Children: Its Clinical Presentation and Characteristic Patterns*, 12 CHILD ABUSE & NEGLECT 163, 166 (1988) (studying 37 sexually abused children, 35 physically abused children, and 130 nonabused children, all of whom were seen at a community mental health clinic, and finding that "[i]nappropriate sexual behavior was noted in 41% of the sexually abused groups, but in less than 5% of the other two groups"); David B. Goldston et al., *Presenting Problems of Sexually Abused Girls Receiving Psychiatric Services*, 98 J. ABNORMAL PSYCHOL. 314, 316 (1989) (studying 128 abused and 67 nonabused girls, all of whom were receiving mental health services and finding that "[i]n a clinic population, the contribution of sexual abuse to psychiatric symptoms is quite specific").

⁹² See MICHAEL S. KRAMER, CLINICAL EPIDEMIOLOGY AND BIostatISTICS 53 (1988).

⁹³ This problem arises even when the symptom is not explicitly used in evaluation as a criterion for distinguishing between abused and nonabused children, but is a means by which children come to the attention of individuals who are interested in diagnosing abuse. For example, if a child exhibits unusual behavior, he or she is more likely to be questioned about abuse. This questioning, in turn, increases the chances that abuse will be discovered. See William N. Friedrich & Redmond A. Reams, *Course of Psychological Symptoms in Sexually Abused Young Children*, 24 PSYCHOTHERAPY 160, 163 (1987) ("For several of the

tingnishing between abused and nonabused children, physical signs will necessarily appear to be more common among abused children than among nonabused children.⁹⁴

A second form of "detection bias" occurs when the diagnosis of abuse increases the chance that a symptom will be detected. This problem arises when children who have been classified as abused are questioned about particular symptoms more vigorously than children who have not been classified as abused. For example, psychologists often probe for behavioral signs of abuse in children who have been diagnosed as abused, but search less vigilantly for behavioral signs in children who have been classified as nonabused.⁹⁵ Moreover, a parent whose child has been diagnosed as abused may selectively recall or construe the child's behavior as evincing the symptom.⁹⁶ As a result, studies often find behavioral signs in abused children at a higher rate than among nonabused children. What remains unknown is whether an equally vigilant search for behavioral signs among nonabused children would have produced an increase in the reported proportion of nonabused children with these symptoms.

children, it was their behavioral deterioration that prompted parental concern, leading to their discovery of abuse."); Teena Sorensen & Barbara Snow, *How Children Tell: The Process of Disclosure in Child Sexual Abuse*, 70 CHILD WELFARE 3, 8 (1991) (finding that for 29% of the children between three to five years old, the impetus for discovery of sexual abuse was "behavioral sign" or "sexualized behavior"). Physical signs of sexual abuse may create this problem as well. See, e.g., S.B. v. P.G.B., 600 So. 2d 256, 257 (Ala. Civ. App. 1991) ("The mother . . . testified that she became concerned that her son had been sexually abused when she noticed redness around his anus following his return from a visit with his father."). "Detection bias" is particularly troublesome in the case of symptoms that are easily visible and/or widely believed to be associated with abuse, because these types of symptoms most often lead to further examination of the child.

⁹⁴ One way to control for "detection bias" in research is to include within the group designated as abused only those children whose abuse status was determined on grounds other than the symptom in question. For example, one could examine the behavior of children whose abuse status was determined by means of medical evidence or through statements by the child, and not with reference to the child's behavior. By testing the ability of the Child Sexual Behavioral Inventory (CSBI) to discriminate between abused and nonabused children using a sample of abused children whose abuse status was not confirmed by means of the CSBI, Friedrich et al. take one step towards this safeguard. Friedrich et al., *supra* note 54, at 304. However, the authors did not control for the possibility that the clinicians in the study relied on less formal reports of sexualized behavior in concluding that abuse occurred.

⁹⁵ If the method of the study is a retrospective review of children's records this problem is especially serious, since the records are limited to those observations that the treating therapist both discovered and thought important for the child's diagnosis. See, e.g., Goldston et al., *supra* note 91, at 315 ("For coding purposes, the absence of a reference to a behavior problem was assumed to reflect an absence of that problem.").

⁹⁶ See Friedrich, *supra* note 91, at 18 (studying children under 35 months of age using the Child Sexual Behavior Inventory, and finding that children involved in custody disputes who the researchers concluded had not in fact been abused received some of the highest scores).

In sum, both types of "detection bias" tend to inflate the value of the relevance ratio. The first type inflates the ratio by pushing the numerator of the ratio toward one (i.e., by increasing the proportion of abused who reportedly exhibit the symptom), and the second type inflates the ratio by pushing the denominator toward zero (i.e., by decreasing the proportion of nonabused who reportedly exhibit the symptom).

Empirical research that compares the symptoms of abused children with those of nonabused children is necessary to identify the probative value of these symptoms for the detection of abuse. However, the research must be carefully examined to ensure that the conclusions it yields about the apparent relevance of symptoms are not tainted by biased research procedures. Apparent relevance may be inflated when the abused sample is drawn from a clinical sample, when the symptom increases the chances that abuse will be detected, or when the diagnosis of abuse increases the likelihood that the symptom will be detected. Because these biases are difficult to eliminate, as a practical matter, courts should treat as presumptively irrelevant symptoms that are only slightly more common among abused children than among nonabused children.

VIII

IS THE ABSENCE OF SYMPTOMS RELEVANT FOR DISPROVING ABUSE?

Relevance or probative value may be tied to the absence as well as to the presence of particular symptoms or behaviors. Just as a finding of vaginal trauma in a young girl may be evidence that penetration occurred, a failure to find such trauma may be evidence that the penetration did not occur. This observation prompts the following questions: If the presence of a symptom is relevant for proving that abuse occurred, is the absence of the symptom relevant for proving that abuse did not occur? And if so, are presence and absence of the symptom equally relevant?

The answer to the first question is "yes": If a symptom is relevant for proving that abuse occurred, the absence of the symptom is also relevant for proving that abuse did not occur.⁹⁷ The answer to the second question is "no": In most instances, presence and absence are not *equally* probative evidence. Indeed, situations exist in which the probative asymmetry between the presence of a symptom and its absence is so great that a court may reasonably find the symptom admis-

⁹⁷ A relevance ratio analysis shows that for mutually exclusive and exhaustive hypotheses, H_0 and H_1 , if evidence E is probative of H_0 , then the absence of E is probative of H_1 .

sible when present, but insufficiently relevant when absent (and vice versa).

To understand why if evidence is relevant for proving that abuse occurred, the absence of such evidence must be relevant for proving that abuse did not occur, assume that (a) children can be classified as suffering from anxiety or as being anxiety-free and (b) anxiety is relevant for proving that abuse occurred. Assumption (b) requires that the percentage of abused children with anxiety is higher than the percentage of nonabused children with anxiety. Therefore, the percentage of nonabused children who are anxiety-free must also be higher than the percentage of abused children who are anxiety-free. Consequently, being anxiety-free is evidence that one was not abused.⁹⁸

Questions about whether the absence of a symptom constitutes relevant evidence arise most frequently in cases involving rape trauma syndrome. Unfortunately, some courts have not reasoned well about the probative value of this evidence. For example, the Kansas Supreme Court argued that "[t]here are no statistics to show that there is any value to a negative finding that the rape trauma syndrome is not exhibited by the alleged victim."⁹⁹ But as indicated above, if statistics demonstrate the relevance of rape trauma syndrome for proving that abuse occurred, those same statistics demonstrate the relevance of a failure to exhibit rape trauma syndrome for proving that abuse did not occur.

The Montana Supreme Court also rejected symptom-absent testimony on the ground that "[t]he absence of evidence of psychological trauma logically does not prove that the offense did not occur."¹⁰⁰

⁹⁸ This may be proved as follows: The relevance ratio associated with evidence E is

$$\frac{P(E|H)}{P(E|\bar{H})} = \frac{X}{Y},$$

and the relevance ratio associated with \bar{E} is

$$\frac{P(\bar{E}|H)}{P(\bar{E}|\bar{H})} = \frac{1-X}{1-Y}.$$

If E is relevant for proving H, then $X > Y$, (because when and only when $X > Y$ will the relevance ratio be greater than one). Furthermore, if $X > Y$ then $-X < -Y$, $1-X < 1-Y$, $1-X/1-Y < 1$. Because the relevance ratio associated with \bar{E}

$$\left(\frac{P(\bar{E}|H)}{P(\bar{E}|\bar{H})} = \frac{1-X}{1-Y} \right)$$

< 1 , \bar{E} is probative of \bar{H} . In general, then, the absence of a symptom is relevant for proving that a condition did not occur whenever the presence of the symptom is relevant for proving that a condition did occur (and vice versa).

⁹⁹ State v. McQuillen, 689 P.2d 822, 830 (Kan. 1984).

¹⁰⁰ State v. Scheffelman, 820 P.2d 1293, 1299 (Mont. 1991) (citing State v. Gilpin, 756 P.2d 445, 451 (Mont. 1988)). See also Myers et al., *supra* note 82, at 64 ("Absence of behaviors does not disprove abuse, but presence of behaviors increases the likelihood of abuse."); Nicole Rosenberg Economou, Note, *Defense Expert Testimony on Rape Trauma Syn-*

Although a lack of trauma does not logically or conclusively prove that rape did not occur, a logical proof test of relevance is too restrictive. Rape trauma syndrome does not conclusively prove that rape occurred, but presence of the syndrome is relevant if it increases the likelihood that rape occurred. Similarly, the lack of trauma may be relevant for proving that abuse did not occur, even though absence of the syndrome does not constitute conclusive proof.

A. Probative Asymmetry

Although the absence of a symptom is relevant if the presence of that symptom is relevant, the relevance of presence or absence for proving or disproving abuse, respectively, is rarely the same. Depending upon the relative proportions of abused and nonabused children who exhibit a symptom, a symptom may be strong evidence of abuse when present, but weak evidence against abuse when absent. Conversely, a symptom may be weak evidence of abuse when present, but strong evidence against abuse when absent.

B. Presence More Relevant Than Absence

When is a symptom strong evidence of abuse when present, but weak evidence against abuse when absent? Generally, this occurs when the evidence is uncommon among abused children, but even more uncommon among children who have not been abused. Specifically, if the sum of the proportion of abused children who exhibit the symptom plus the proportion of nonabused children who exhibit the symptom is less than 100%, then the presence of the symptom is more probative for proving abuse than is the absence of the symptom for disproving abuse.¹⁰¹

As noted above, this is true of sexually transmitted diseases ("STD's"), which occur in about 5% of abused children, but almost never occur in nonabused children. Although the presence of an STD is strong evidence that a child was abused, the absence of an STD is weak evidence that a child was not abused. An example illustrates this point. Suppose that 5% of sexually abused children are symptomatic and only 1% of nonabused children are symptomatic. The presence of an STD is moderately relevant, because it yields a relevance ratio of $.05/.01 = 5$. How relevant is the absence of the symptom for

drome: Implications for the Stoic Victim, 42 HASTINGS L.J. 1143, 1155 (1991) ("Although evidence of the degree to which a woman suffered physical or emotional injuries does tend to corroborate a rape charge, a lack of perceptible injuries or trauma does not demonstrate that a woman was not raped.")

¹⁰¹ As the probative value of the symptom's presence increases (i.e., as the ratio of the two proportions increases), the relevance of the symptom's absence also increases, but not as steeply.

disproving abuse? The relevance ratio for proof that abuse did not occur requires one to compare the proportion of asymptomatic abused children and the proportion of asymptomatic nonabused children.¹⁰² Because 5% of abused children are symptomatic, 95% of abused children are asymptomatic. Similarly, 1% of nonabused children are symptomatic, 99% of nonabused children are asymptomatic. Thus, the absence of the symptom is only slightly relevant, because it yields a relevance ratio of $.99/.95 = 1.04$. Although relevance ratios that exceed one are, technically speaking, relevant, relevance ratios that are so close to one are probably best excluded on grounds of minimal relevance.¹⁰³

C. Absence More Relevant Than Presence

When is a symptom weak evidence of abuse when present, but strong evidence against abuse when absent? Generally, this occurs when the evidence is common among abused children, and only slightly less common among nonabused children. Specifically, if the sum of the proportion of abused children who exhibit the symptom plus the proportion of nonabused children who exhibit the symptom is more than 100%, then the presence of the symptom is less probative for proving abuse than is the absence of the symptom for disproving abuse.

This analysis highlights the problems associated with developing profiles of abused children that describe as many abused children as possible. Profiles that describe a majority of abused children are also likely to describe a large percentage of nonabused children. In such situations, evidence of a failure to fit the profile may provide a strong basis for believing that the child was not abused, whereas evidence of a match may provide only a weak basis for believing that a child was abused.¹⁰⁴

The probative asymmetries discussed above are quite common, as an everyday example demonstrates. For instance, whereas the presence of a computer in Jim's office is not particularly diagnostic of the hypothesis that Jim is a scientist, the absence of a computer in Jim's office is diagnostic of the hypothesis that Jim is not a scientist. From a

¹⁰² See *supra* note 97.

¹⁰³ See *supra* Part VII.

¹⁰⁴ For example, the Child Sexual Behavior Inventory is a measure of the range and frequency of a child's sexual behavior. One study which employed statistically derived cut-off scores reported that the Inventory correctly classified 92% of abused 2-6 year old boys as abused, and incorrectly classified 38% of nonabused boys as abused. Friedrich et al., *supra* note 54, at 309. Although scores above the cut-off are relevant for proving abuse (relevance ratio = $.92/.38 = 2.4$), scores below the cut-off are stronger evidence that the child was not abused (relevance ratio = $.62/.08 = 7.8$). As discussed *supra* note 54, however, a number of specific behaviors measured by the Inventory are highly probative of abuse.

relevance ratio perspective, the presence of an office computer is common to both scientists (e.g., $p=.95$) and nonscientists (e.g., $p=.70$). Hence, the relevance ratio will be near one. On the other hand, because the absence of a computer among scientists is an infrequent event (e.g., $p=.05$) relative to the absence of a computer among nonscientists (e.g., $p=.30$), the relevance ratio associated with the absence of a computer is six to one.

D. Implications of Asymmetry for Parties' Discovery Rights

That symptom relevance varies depending upon whether the symptoms are present or absent has implications for assessing the parties' rights to collect such evidence. For example, defendants sometimes seek the court-ordered examination of an alleged victim on the ground that such an examination may yield relevant evidence for the defendant's defense. Prosecutors often resist such a request on the ground that the alleged victim's privacy rights outweigh the probative value of whatever evidence such an examination may yield. Judicial resolution of this tradeoff should depend, in part, on a proper understanding of the probative value of the evidence that one may discover. Thus, in cases where a symptom's presence is more relevant than a symptom's absence, it would not necessarily be inconsistent for a judge to require an examination of the victim when requested by the prosecution, but to deny such a request when made by the defense.¹⁰⁵

IX

DOES THE PRESENCE OF SYMPTOMS AMONG THE NONABUSED MAKE THESE SYMPTOMS IRRELEVANT?

We have heretofore emphasized the problems inherent in undue focus on the numerator of the relevance ratio (i.e., the proportion of abused children who exhibit a symptom). Thus, one might conclude that all of the problems with scientific evidence would be solved if commentators and courts would simply pay more attention to the denominator (i.e., the proportion of nonabused children who exhibit the symptom). However, overemphasizing the likelihood that nonabused children will exhibit a symptom is equally possible (i.e., focusing too much on the denominator). An undue focus on either the numerator or denominator misses the importance of assessing the

¹⁰⁵ Although the defendant did not raise the issue of a court-ordered examination, the Indiana Supreme Court has held that a defendant must be allowed to introduce expert testimony that an alleged victim's behavior is inconsistent with that of a rape victim, *Henson v. State*, 535 N.E.2d 1189, 1193 (Ind. 1989), reasoning that "this Court has already recognized the admissibility of rape trauma syndrome evidence It would be fundamentally unfair to allow the use of such testimony by the State . . . and then deny its use by a defendant" *Id.*

relation between the two in determining whether a symptom is relevant for proving that abuse occurred.

Critics of expert testimony frequently argue both that symptoms of abuse are uncommon and that symptoms occur among nonabused children, thereby mentioning both numerator and denominator of the relevance ratio. The critics present these arguments piecemeal, however, as if each argument alone were sufficient to render symptoms irrelevant.¹⁰⁶ When critics discuss the need for a direct comparison between abused and nonabused children, they treat it as a methodological criticism of the research on sexual abuse, rather than as the fundamental issue in assessing the relevance of sexual abuse symptoms.¹⁰⁷

Even when the courts do not ignore the denominator, they sometimes exaggerate its importance. In *Dunkle*, the Pennsylvania Supreme Court asserted that to discriminate between abused and nonabused children, symptoms must not only "appear regularly in a group of children with a certain experience, but . . . must not appear in other groups of children who have not had that experience."¹⁰⁸ The court glossed over the distinction made by the authority it cites, which argues that the presence of certain symptoms may in fact "increase the likelihood," but not "clearly indicate," that sexual abuse has

¹⁰⁶ See, e.g., *Commonwealth v. Dunkle*, 602 A.2d 830, 835 n.16 (Pa. 1992) (discussing the existence of symptoms among nonabused children, but treating statistics regarding numerator as a sufficient basis for concluding that "these types of percentages cannot constitute probative evidence."); *State v. Foret*, 628 So. 2d 1116, 1127 (La. 1993) ("Criticisms include the varying reactions children have to abuse [a numerator problem] and the fact that behavior often attributed to abuse is sometimes the result of other emotional problems that do not stem from abuse [a denominator problem]."); Askowitz, *supra* note 62, at 209 ("These studies have revealed the absence of a single, typical child sexual abuse victim In addition, certain behaviors may seem to indicate some other traumatic event or behavior of normal children.").

¹⁰⁷ See Askowitz, *supra* note 62, at 209 ("Methodological weaknesses in the existing studies of child sexual abuse also contribute to the questionable reliability of testimony Without comparison data, it is impossible to conclude that observed effects significantly differentiate abused from non-abused children.") (emphasis added). *State v. Reser*, 767 P.2d 1277 (Kan. 1989) is a good example of courts' sometimes limited understanding of the importance of the comparisons. In *Reser*, the Kansas Supreme Court acknowledged that "comparison testing" had not been performed on the behavioral symptoms of sexual abuse, but, nevertheless, allowed expert testimony on behavioral symptoms. *Id.* at 1282. The court reasoned that because the expert at trial had "merely testified as to characteristics common to children who had been sexually abused and gave her opinion that the victim in this case exhibited some of these characteristics," the evidence should be allowed. *Id.* The expert had not expressed an ultimate opinion that the child had been abused based on the existence of common characteristics. *Id.* What the court in *Reser* failed to acknowledge, or perhaps recognize, was that a lack of comparison testing makes it impossible to determine whether common characteristics increase the likelihood that abuse occurred, let alone to understand whether they justify a conclusion that abuse was likely.

¹⁰⁸ *Dunkle*, 602 A.2d at 832 n.2 (quoting JEFFREY J. HAUGAARD & N. DICKON REPPUCI, *THE SEXUAL ABUSE OF CHILDREN: A COMPREHENSIVE GUIDE TO CURRENT KNOWLEDGE AND INTERVENTION STRATEGIES* 177-78 (1988)).

occurred.¹⁰⁹ If they occur among a higher percentage of abused children, symptoms can occur among some nonabused children and still be relevant evidence that abuse occurred.¹¹⁰

Although we are critical of a single-minded focus on the denominator in assessing the relevance of symptoms, we are sympathetic to the argument that if a large proportion of nonabused children exhibit a certain symptom, then the symptom should not be considered relevant unless there is *convincing* evidence that the symptom is more common among abused children.¹¹¹ There are several reasons why this is so. First, if symptoms are common among nonabused children, then the relevance ratio cannot be terribly high, no matter how common the symptoms are among abused children. For example, if 50% of nonabused children exhibit a symptom, then the relevance ratio cannot be higher than two because no more than 100% of abused children will exhibit the symptom. A high denominator thus puts a serious limit on the possible size of the relevance ratio.¹¹²

Second, a high denominator is unlikely to be the result of "selection bias." As discussed earlier, "selection bias" occurs when abused subjects are drawn from a clinical population, inflating the number of abused children exhibiting a symptom. Sexually abused children are difficult to identify, if one does not recruit through clinics treating abused children. On the other hand, nonabused children are relatively easy to locate, making it possible to draw samples representative of the general population.¹¹³ Therefore, whereas a high numerator in a study comparing abused and nonabused children might be due to the source from which the sample of abused children was drawn, a high denominator is more likely attributable to a high occurrence of the symptom among nonabused children in general.¹¹⁴

¹⁰⁹ HAUGAARD & REPPUCCI, *supra* note 108, at 178.

¹¹⁰ If a symptom is rendered irrelevant by the possibility that a nonabused child exhibits such a symptom, then a child's sworn testimony that he or she was abused would be irrelevant, because some nonabused children either lie or falsely believe that they were abused.

¹¹¹ Some medical experts have adopted the rule of thumb that if a symptom appears among more than 10% of nonabused children, the symptom is considered a "normal" finding. See Adams, *supra* note 37; Berenson et al., *supra* note 35.

¹¹² In contrast, note that a low numerator fails to limit the size of the relevance ratio. Even if only 1% of abused children exhibit a symptom, the relevance ratio could be infinitely high, because as few as 0% of nonabused children might exhibit the symptom.

¹¹³ For example, in the Friedrich et al. study, *supra* note 54, the authors drew the abused sample from clinics, leading to "selection bias" problems, whereas they drew the nonabused sample from children visiting their pediatrician for well-child checkups.

¹¹⁴ Although researchers will exclude from the nonabused sample children about whom suspicions of abuse have arisen, see, e.g., Friedrich et al., *supra* note 54, some nonabused children might still exhibit the symptom, because they experienced abuse that has never been detected. Even if this occurred, however, the percentage of nonabused children in general who exhibit the symptom should not be large, because (1) most children have not been abused, (2) many children who have been abused have been identi-

Although the most egregious errors in assessing relevance occur when one focuses exclusively on the numerator of the relevance ratio, an undue focus on the denominator can be equally misleading. That some nonabused children exhibit symptoms can lead one to prematurely dismiss such symptoms as irrelevant for proving that abuse occurred. Relevance ratio analysis requires that one determine the proportion of abused children who exhibit a symptom, and compare this proportion to the proportion of nonabused children who exhibit the symptom. When the percentage of nonabused children who exhibit the symptom is high, the relevance ratio counsels skepticism; in such a case, admissibility should be contingent on clear evidence that a higher proportion of abused children are symptomatic.

CONCLUSION

The relevance ratio is simple, analytically compelling, and easy to apply. It not only indicates whether proffered evidence is probative, but also measures how probative the evidence is. When courts examine expert testimony through the relevance ratio lens, they can spot errors in reasoning quickly and avert them. The relevance ratio also facilitates the resolution of problems that, upon first impression, seem more vexing. These include the relevance of symptoms for rebuttal purposes and the relative probative value of the presence and absence of symptoms. In short, use of the relevance ratio by courts would lead to admissibility decisions that are readily defensible on scientific and logical grounds, and an ability to evaluate the appropriate weight that should be accorded to certain expert testimony.

Although trial judges in the post-*Daubert* era must think more like scientists when evaluating the probative value of evidence, they must also continue to act like judges and make decisions in accordance with the Rules of Evidence. Because the relevance ratio builds upon the legal definition of relevance embodied in Federal Rule of Evidence 401, it promotes both goals. This fusion of science and legal doctrine is healthy and necessary.

Our focus has been the heuristic value of the relevance ratio for clarifying and understanding the probative value of expert testimony in child sexual abuse cases, particularly testimony that invokes the ambiguous phrase "consistent with" abuse and "common among" abused

fied, and (3) of those who were abused, but who have not been identified, only a fraction will exhibit the symptoms in question. For example, suppose that 30% of children are abused before they are 18 and that in 90% of those cases, the abuse is never detected. If this were the case, then only 1% of the nonabused population would have been sexually abused in the past year without anyone knowing. If a symptom of abuse, in fact, appears among 30% of abused children and 1% of nonabused children, then a study of children believed to be nonabused will find that 1.3% of the nonabused children exhibit the symptom, only slightly underestimating the probativeness of the symptom.

children. The relevance ratio is also useful for thinking about expert testimony in other types of cases, particularly those that include forensic science testimony. For example, consider the testimony of ballistics expert William H. Proctor in the Sacco and Vanzetti murder case.¹¹⁵ Proctor had examined the Colt bullet found in the body of one of the victims and the Colt automatic found in Sacco's possession. He testified that the bullet was "consistent with being fired by that pistol."¹¹⁶ Was Proctor's opinion relevant evidence that the bullet did in fact come from Sacco's gun? One may wish the defense or the judge had been familiar with the relevance ratio.¹¹⁷ Subsequently, Proctor acknowledged that "[h]ad I been asked the direct question: whether I had found any affirmative evidence whatever that this so-called mortal bullet had passed through this particular Sacco's pistol, I should have answered then, as I do now without hesitation, in the negative."¹¹⁸

¹¹⁵ Commonwealth v. Sacco, 151 N.E. 839 (Mass. 1926).

¹¹⁶ *Id.* at 862.

¹¹⁷ In his charge to the jury, the judge told the jurors that Proctor had testified that "it was his [Sacco's] pistol that fired the bullet that caused the death of Beradelli." *Id.* at 863.

¹¹⁸ *Id.* at 862.