Einstein on the Bench?: Exposing What Judges Do Not Know About Science and Using Child Abuse Cases to Improve How Courts Evaluate Scientific Evidence

JOËLLE ANNE MORENO*

It has been a decade since the Supreme Court made judges the arbiters of scientific validity through Daubert v. Merrell Dow Pharmaceuticals, Inc.** *Although this decision was intended to improve how courts use science, recent empirical evidence reveals that judges continue to struggle with scientific evidence and that* Daubert *has failed to yield accurate or consistent decisions. This also means that judges have received little useful guidance from ten years of academic literature expounding on the science-law chasm.*

If the academic discourse is not helpful, it may be because non-scientists too often try to tame science by treating it as a single discipline, which strips away context and meaning. This article takes a different approach. It explores the admissibility of complex medical evidence offered to defeat allegations of child abuse. These cases offer a useful theoretical model of the interaction of science and law because the fact that children can only be injured by accident or abuse limits confounding factors, such as numerous potentially valid diagnoses and the unpredictable influence of plaintiff/victim testimony. In practice, better judicial decision-making in child abuse cases can improve or even save children's lives. Approximately one million children are abused and/or neglected every year in the United States. In 1997, twenty-five percent of child abuse-related homicides occurred after state investigations had concluded that it was safe to return the child to her home. Child abuse cases illustrate why an accurate understanding of science is vital to law, because the life of a child— and not some abstract principle or legal theory— may hang in the balance.

I. INTRODUCTION

In all cases involving empirical evidence, good law depends on good science. Judges and juries cannot make accurate legal decisions if invalid scientific evidence distorts their understanding of the facts. Recently, it has become popular

---

* Assistant Professor of Law, New England School of Law. I am indebted to my colleagues, especially Professor Peter M. Manus for his hard work on a related conference, Christina Shea for her brilliant insight into the structure of my ideas, and Professor David M. Siegel for his enthusiastic support and incisive critique. I would like to thank Dean John F. O'Brien and the Trustees of New England School of Law for funding this research through a James R. Lawton Research Grant. Finally, I gratefully acknowledge the essential role played by numerous forensic pediatricians and pediatric radiologists who have dedicated their careers to detecting, treating, and preventing abuse.

** 509 U.S. 579 (1993).

sport for scientists to blame judges, lawyers, or jurors for legal decisions that misapply basic principles of science.[1] Although *Daubert v. Merrell Dow Pharmaceuticals, Inc.*[2] was intended to address the growing concern about junk science in the courts, the creation of a judicial gatekeeping role for scientific evidence did fix the problem. And although the alleged disjuncture between science and law continues to be fertile scholarly terrain,[3] the academic discourse often ignores the practical problems faced by judges, lawyers, and jurors.[4] The *Daubert* Court intended to radically transform the functional, rather than theoretical, relationship between science and law by forcing judges to play a new, more active role in enhancing the quality of scientific evidence used to decide legal cases.[5] According to Justice Breyer, judges must do much more than simply

---

[1] *See generally*, KENNETH R. FOSTER & PETER W. HUBER, JUDGING SCIENCE: SCIENTIFIC KNOWLEDGE AND THE FEDERAL COURTS 17 (1999) (describing junk science as a legal, rather than scientific, problem "cultivated by the adversarial nature of legal proceedings"); MARCIA ANGELL, SCIENCE ON TRIAL: THE CLASH OF MEDICAL EVIDENCE AND THE LAW IN THE BREAST IMPLANT CASE (1996) (describing how pseudoscience influences the outcome of legal cases).

[2] 509 U.S. 579, 589 (1993) (establishing that the trial judge must ensure that all scientific testimony or admitted evidence is not only relevant but reliable).

[3] *See, e.g.*, Jan Beyea & Daniel Berger, *Scientific Misconceptions Among* Daubert *Gatekeepers: The Need for Reform of Expert Review Procedures*, 64 LAW & CONTEMP. PROBS. 327 (2001); Daniel J. Capra, *The* Daubert *Puzzle*, 32 GA. L. REV. 699 (1998); David L. Faigman, *Appellate Review of Scientific Evidence Under* Daubert *and* Joiner, 48 HASTINGS L.J. 969 (1997); Michael H. Gottesman, *Admissibility of Expert Testimony After* Daubert*: The "Prestige" Factor*, 43 EMORY L.J. 867 (1994); Jay P. Kesan, *An Autopsy of Scientific Evidence in a Post-*Daubert *World*, 84 GEO. L.J. 1985 (1996); Derek L. Mogck, Note, *Are We There Yet?: Refining the Test for Expert Testimony Through* Daubert, Kumho Tire *and Proposed Federal Rule of Evidence 702*, 33 CONN. L. REV. 303 (2000).

[4] Science impacts legal decisions in a wide range of cases. Mary Sue Henifin et al., *Reference Guide on Medical Testimony*, *in* REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 439, 441 (Fed. Judicial Ctr. ed., 2000) ("Testimony by physicians is one of the most common forms of expert testimony in the courtroom today. Medical testimony is routinely offered in both civil and criminal cases . . . .") [hereinafter REFERENCE MANUAL]. The *Reference Manual* was first published in 1994 as a response to the Supreme Court's decision in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993). Although it provides a wide range of helpful and practical information on science and law, it should be noted that the *Reference Manual* has been criticized as defense oriented by various plaintiffs' organizations. *See, e.g.*, Joseph T. Walsh, *Keeping the Gate: The Evolving Role of the Judiciary in Admitting Scientific Evidence*, 83 JUDICATURE 140 (1999).

[5] This effort to change legal practice was codified to conform to the developing Supreme Court doctrine in December 2000, through extensive amendments to Federal Rule of Evidence 702, Testimony by Experts:

> If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise, *if (1) the testimony is based upon sufficient facts or data, (2) the*

reject specious science; they must "aim for decisions that, *roughly* speaking, *approximately* reflect the scientific 'state of the art.' "[6]

It should come as no surprise that many judges, speaking candidly, consider themselves ill-suited to the task of evaluating scientific evidence and could not begin to articulate the scientific state of the art.[7] In fact, recent empirical evidence shows that most judges cannot explain even the most common principles of scientific methodology. In October 2001, the first comprehensive national study assessing the scientific acumen of 400 state court judges was published.[8] This study, which focused on how judges use the *Daubert* criteria to make legal decisions about scientific evidence,[9] revealed that although "judges overwhelmingly support the [*Daubert*] 'gate-keeping' role, . . . many of the judges surveyed lacked the scientific literacy seemingly necessitated by *Daubert*."[10] In fact, 96% of the judges failed to demonstrate even a basic understanding of two of the four *Daubert* criteria.[11] This means that, a decade after *Daubert*, courts have systemic and ongoing problems assessing the quality of scientific evidence. It is difficult to reconcile such staggering levels of scientific ignorance with the increasing importance of science and technology to society and to law.[12]

The recent proliferation of academic literature expounding on the interdisciplinary chasm has done little to educate judges who must grapple with

---

testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case.

FED. R. EVID. 702 (italics indicate amendments to the previous rule).

[6] Justice Stephen J. Breyer, The Interdependence of Science and Law, Address for the Association for the Advancement of Science Annual Meeting and Science Innovation Exposition (Feb. 16, 1998) (transcript available at http://aaas.org/meetings/1998/breyer98.htm).

[7] In a recent national survey of four hundred state court judges, 48% stated that they were not adequately prepared to deal with the range of scientific evidence proffered in their courtrooms. Sophia I. Gatowski et al., *Asking the Gatekeepers: A National Survey of Judges on Judging Expert Evidence in a Post-*Daubert *World*, 25 LAW & HUM. BEHAV. 433, 442 (2001).

[8] *See id.* at 433–35.

[9] *See* Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 593–94 (1993) (identifying as flexible guidelines for the court: (1) testability; (2) peer review and publication; (3) error rates; and (4) general acceptance).

[10] Gatowski, *supra* note 7, at 433.

[11] *See id.* at 444–47 (detailing empirical evidence that shows that judges do not understand the *Daubert* criteria of (1) testability/falsifiability and (2) error rates).

[12] In light of "the significant advances in science and technology in the twentieth century[,] . . . . [a] substantial level of sophistication in the scientific method will be necessary if judges are ever going to integrate science successfully into their legal decisions." David L. Faigman, *Mapping the Labyrinth of Scientific Evidence*, 46 HASTINGS L.J. 555, 560–79 (1995).

real scientific evidence in actual cases.[13] This may be attributable in part to the tendency of non-scientists to treat all science as if it were a single discipline.

> The notion that there is a simple, identifiable, universal scientific method used in some kind of standard way by scientists to distinguish science from non-science is difficult to support on any kind of empirical basis. One of the factors which illustrates the implausibility of this contention is the sheer diversity of activities which can be placed beneath the umbrella of modern science.[14]

If judges cannot divine specific guidance from generalized discussion of science and law, we need a new approach.[15] As a first step, we should avoid the temptation to treat all science as a single field, which strips away meaning and practical value. To obtain greater insight into scientific processes and develop a viable model for judges to analyze scientific testimony, the article will explore competing medical diagnoses offered in child abuse cases.

Child abuse cases offer a useful and enlightening model of the interaction between science and law. As in all cases with a scientific determinant, causation is the threshold legal question. Although the medical evidence in child abuse cases may be complex, these cases almost always involve a binary decision—were the injuries caused by abuse or accident? Most child abuse victims are infants or toddlers.[16] Very young children only sustain certain types of injuries,

---

[13] *See* Gatowski et al., *supra* note 7, at 456–58 (citing thirty articles on science and law published between 1994 and 1999).

[14] Gary Edmond & David Mercer, *Trashing Junk Science*, 1998 STAN. TECH. L. REV. 3, 29, *available at* http://stlr.stanford.edu/STLR/Articles/98_STLR_3.

[15] This conclusion is consistent with the findings of Professors Ronald J. Allen and Ross M. Rosenberg who recently explored the impact of legal theory on the legal process. *See* Ronald J. Allen & Ross M. Rosenberg, *Legal Phenomena, Knowledge, and Theory: A Cautionary Tale of Hedgehogs and Foxes*, 77 CHI.-KENT L. REV. 683 (2002). They began with the hypothesis that "three variables—ambiguity, unpredictability and common sense reasoning—determine to a significant extent the explanatory power and usefulness of top-down generalized theories to legal phenomena." *Id.* at 687. They tested this hypothesis by searching the relationship between citations in law reviews to various renowned theoreticians and citations in cases or legislative histories. They found that courts and legislators tend to ignore theorists and cite practitioners who offer more specific and useful guidance. *Id.* at 693. This leads to the conclusion that "judges apparently, and not surprisingly, are looking for answers to discrete questions, not solutions grounded in grand theory." *Id.*

[16] *See* Thomas D. Lyon et al., *Child Abuse: Medical Evidence of Physical Abuse in Infants and Young Children*, 28 PAC. L.J. 93, 101 (1996) ("Children under eighteen months of age suffer 80% of the fractures attributable to child abuse."); U.S. DEP'T OF HEALTH & HUMAN SERVS., NATIONAL CHILD ABUSE AND NEGLECT DATA SYSTEM, SUMMARY OF KEY FINDINGS FROM CALENDAR YEAR 2000, *at* http://www.calib.com/nccanch/pubs/factsheets/canstats.cfm (last updated Feb. 24, 2003) (indicating that the victimization rate for children younger than three years old was "15.7 victims per 1,000" while the rate "for children ages 16 and 17 was 5.7 victims per 1,000") [hereinafter SUMMARY 2000].

such as broken bones, by accident or abuse. With just two possible diagnoses, the *only* scientific evidence relevant to causation is evidence that makes one of these two causal explanations more or less probable.[17] In child abuse cases involving children too young to testify, medical science will be offered by both parties to explain the etiology of the injuries. Accident histories will be tested against a standard of medical plausibility. Only after science has been used to determine causation will the law proceed to identify the perpetrator and impose civil or criminal sanctions.[18] An exploration of medical science in this context provides two additional benefits.

First, the child abuse case model explains how courts should deal with complex, novel, and controversial scientific evidence. The most frequent medical defense to child abuse cases is that the child suffers from Osteogenesis Imperfecta ("OI"), which is also known as "Brittle Bone Disease."[19] There is currently a global debate focused on the scientific validity of medical expert testimony diagnosing variant forms of OI.[20] In these cases, a small but growing number of medical experts have been permitted to testify for the defense that fractures sustained by very young children are not indicative of abuse, but are instead a natural and unintended complication of certain rare and transient metabolic bone diseases, such as Temporary Brittle Bone Disease ("TBBD"). As of October 2000, a single medical expert had provided expert testimony diagnosing TBBD in 103 child abuse cases. In seventy-eight of these cases, judges admitted the

---

[17] *See* FED. R. EVID. 401 (" 'Relevant evidence' means evidence having 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.").

[18] *See* REFERENCE MANUAL, *supra* note 4, at 445 (describing how legal rules permit medical experts to testify "on one or more of the ultimate issues in the case, such as causation").

[19] *See infra* notes 150–53 and accompanying text.

[20] Ralph S. Lachman et al., *Differential Diagnosis II: Osteogenesis Imperfecta*, *in* DIAGNOSTIC IMAGING OF CHILD ABUSE 197, 210 (Paul K. Kleinman ed., 2d ed. 1998) (describing how the "advocacy of the concept of transient brittle bone disease in court cases of alleged child abuse, have stirred intense controversy in the United Kingdom as well as in North America").

On April 11, 2001, Judge Peter Singer, of the Family Division of the Royal Courts of Justice England issued a lengthy decision addressing the validity of Temporary Brittle Bone Disease. According to Judge Singer:

> In short, and having considered carefully the way in which those in this case and others in the literature have expressed conclusions against the existence of TBBD as an identifiable disorder, I can only say that in my judgment its existence is very far from proven. It remains at best a highly controversial theory. Unless and until a far broader section of the medical community accepts its existence, I for my part very much doubt whether it can be appropriate for courts in this jurisdiction to have such an as yet unaccepted hypothesis as TBBD presented as an explanation for fractures in children.

Re X (Non-Accidental Injury: Expert Evidence), 2 F.L.R. 1, 19 (Royal Courts of Justice, Fam. Div. 2001).

medical evidence, the defense prevailed at trial, and the children were returned to their homes.[21]

Second, child abuse cases enable us to focus quite explicitly on what Professor Ronald J. Allen has referred to as the "real" question: "how [does] expert testimony fit[] into the administration of justice more generally?"[22] If we can improve the quality of pretrial determinations, this should enhance adjudicative accuracy, and—in the context of OI child abuse cases—we could save lives. There is a powerful imperative to ensure judicial accuracy in child abuse cases. Child abuse transcends all social, political, and economic boundaries. In the United States, more than 879,000 children are abused and/or neglected every year.[23] When child abuse results in death or serious physical injury, physicians, social service agencies, law enforcement, and local prosecutors must coordinate their efforts to serve the medical and legal interests of child victims and the state. Mistakes in child abuse cases are costly and sometimes fatal.[24] Every year more than 1,200 abused children die from their injuries.[25] Many of these deaths may be preventable. In 1997, 226 children who were returned to their homes following official abuse inquiries were later beaten to death.[26]

How judges frame the reliability inquiry can influence how well they understand the scientific questions and improve the accuracy of their legal decisions. In previous work, I have placed myself at the center of a lively debate about whether courts must or should attempt a global comparison of the tenets of a proffered scientific discipline. I have argued that judges might operate more effectively if they confined their analysis to a more manageable assessment of the science necessary to evaluate the relevant facts.[27] Regardless of one's view on this question as a matter of theory, as a practical matter, courts struggle with scientific evidence and might benefit from useful guidance. The four *Daubert*

---

[21] *See Evidence of Disease Led to Return of 78 Children*, HERALD (Glasgow), Oct. 17, 2000, at 3, LEXIS, Nexis Library, GHERLD File.

[22] Ronald J. Allen, Expertise and the Supreme Court: What is the Problem? (March 6, 2003) (unpublished manuscript, on file with the author).

[23] *See* SUMMARY 2000, *supra* note 16.

[24] *See* Deborah S. Ablin & Shashikant M. Sane, *Non-Accidental Injury: Confusion with Temporary Brittle Bone Disease and Mild Osteogenesis Imperfecta*, 27 PEDIATRIC RADIOLOGY 111, 111 (1997) ("To send a child home to the same abusive environment may result in his death or severe morbidity.").

[25] *See* SUMMARY 2000, *supra* note 16.

[26] *See* N. Dickon Repucci & Carrie S. Fried, *Child Abuse and the Law*, 69 UMKC L. Rev. 107, 120 (2000) (citing U.S. DEP'T OF HEALTH & HUMAN SERVS., CHILD MALTREATMENT 1999: REPORTS FROM THE STATES TO THE NATIONAL CHILD ABUSE AND NEGLECT DATA SYSTEM app. E, tbl.E-14 (1999), *available at* http://www.acf.dhhs.gov/programs/cb/publications/cm99 [hereinafter CHILD MALTREATMENT 1999]).

[27] *See generally*, Joëlle Anne Moreno, *Beyond the Polemic Against Junk Science: Navigating the Oceans That Divide Science and Law with Justice Breyer at the Helm*, 81 B.U. L. REV. 1033 (2001).

factors, crafted as a tool to reconcile both local and global reliability, "have too often been leaden deadweights woodenly applied, inert impediments to the development of a sophisticated approach by the courts."[28] A useful judicial inquiry must be "content specific to the case"[29] and should help resolve the question of the appropriate scope of the reliability determination. An effective inquiry must also focus the judge on the question of whether the expert has applied reliable scientific information to the facts and drawn appropriate inferences and conclusions.

This article, which is divided into five parts, uses a real scientific controversy to construct a practical solution. "While there are innumerable specialized fields in science today, and while knowledge in one field does not necessarily transfer to another field, there are, nevertheless, general standards applicable to all fields of science that distinguish genuine science from pseudo-science and quack science."[30] The goals are (1) to create a context-specific model that describes how we use and understand novel, complex medical evidence and (2) to enhance our understanding of validity in cases involving a scientific determinant.

Although child abuse cases are dominated by scientific evidence and have dramatic real world consequences, they have been completely ignored by those who study the interaction between science and law. Part II of this article defines the scope of the problem using recent empirical evidence to expose how poorly judges understand basic science. These data prove that at the state court level, where the overwhelming majority of child abuse cases are decided, admissibility standards are not being accurately or consistently applied to medical expert evidence. Part III briefly describes the legal standard governing the admissibility of scientific evidence in the federal courts and most state courts. Part IV introduces the child abuse model, explaining how science and law are interdependent from the initial detection of suspected abuse, throughout the investigation and adjudication of the legal case. Part V describes medical evidence frequently admitted in child abuse cases. Part VI develops and then applies a new pretrial inquiry, which moves the judge chronologically through a simplified version of the scientific process. Finally, the article concludes with a discussion of the future roles of the judge, lawyer, and scientific expert in improving adjudicative accuracy.

---

[28] Mark P. Denbeaux & D. Michael Risinger, Kumho Tire *and Expert Reliability: How the Question You Ask Gives the Answer You Get*, 34 SETON HALL L. REV. (forthcoming 2003) (manuscript at 18, on file with author).

[29] *Id.* at 19.

[30] Lee Loevinger, *Science and the Legal Rules of Evidence: A Review of Galileo's Revenge: Junk Science in the Courtroom*, 32 JURIMETRICS J. 487, 500 (1992) (reviewing PETER W. HUBER, GALILEO'S REVENGE: JUNK SCIENCE IN THE COURTROOM (1991)).

II. A DECADE AFTER *DAUBERT V. MERRELL DOW PHARMACEUTICALS, INC.*, JUDGES STRUGGLE WITH THEIR GATEKEEPING ROLE

A. Daubert *Changes the Rules*

Daubert dramatically transformed the role of the judge in cases involving scientific evidence.[31] The *Daubert* Court abandoned the long-standing *Frye*[32] inquiry, which had limited the judge's role and used "general acceptance" as a surrogate for scientific validity.[33] Justice Blackmun, writing for the *Daubert* majority, quickly concluded that the *Frye* test had not survived the recent adoption of Rule 702 of the Federal Rules of Evidence.[34] After *Daubert*, a mere finding of general acceptance would not guarantee admission of scientific

---

[31] *See* Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 589 (1993) (establishing the new gate-keeping role for the court).

[32] *See id.* at 588 (citing Frye v. United States, 293 F. 1013 (D.C. Cir. 1923)). In *Frye v. United States*, the federal courts first recognized a special rule governing the admissibility of scientific evidence in 1923. In *Frye*, the D.C. Court of Appeals upheld a district court decision to refuse to admit the results of a systolic blood pressure detection test (a precursor to the polygraph) on the ground that the test had not gained "general acceptance" as a method of assessing truth telling. *See Frye*, 293 F. at 1014. The "general acceptance" standard did not die with *Frye*, but has been incorporated into the *Daubert* analysis. *See Daubert*, 509 U.S. at 594; *see also Standards and Procedures for Determining the Admissibility of Expert Evidence After* Daubert, 157 F.R.D. 571, 572 (1994) (describing how the "general acceptance" standard allows judges to defer to the scientific community and avoid the difficulties of evaluating confusing or technical information outside the court's area of expertise). *Frye* is also the current rule in twenty states. *See infra* note 42 (detailing the different state admissibility standards).

[33] The parties in *Daubert* advanced conflicting arguments about which rule should be applied if the Court abandoned the *Frye* test. *See Daubert*, 509 U.S. at 588. The plaintiffs argued that the Federal Rules of Evidence should be interpreted to admit all relevant testimony proffered by a qualified expert. *See id.* The defendant argued that any new rule fashioned by the Court must evaluate the reliability of the conclusions proffered by the scientific expert. *See* Gottesman, *supra* note 3, at 869.

[34] According to the Court, "a rigid 'general acceptance' standard would be at odds with the 'liberal thrust' of the Federal Rules [of Evidence] and their general approach of relaxing the traditional barriers to 'opinion' testimony." *Daubert*, 509 U.S. at 588 (citations omitted). In addition, by 1993, the Court was aware of the widespread perception that the *Frye* test was used to exclude reliable scientific expert testimony because it related to novel or developing scientific theories or techniques. *See, e.g.*, Michael H. Graham, *The Expert Witness Predicament: Determining "Reliable" Under the Gate-Keeping Test of* Daubert, Kumho, *and Proposed Amended Rule 702 of the Federal Rules of Evidence*, 54 U. MIAMI L. REV. 317, 320 (2000) (noting that, as a practical matter, *Frye* was applied only to new or novel forensic evidence offered by the government in criminal cases); Kristina L. Needham, Note, *Questioning the Admissibility of Nonscientific Testimony After* Daubert*: The Need for Increased Judicial Gatekeeping to Ensure the Reliability of All Expert Testimony*, 25 FORDHAM URB. L.J. 541, 544–45 (1998) (describing how the *Frye* test tended to exclude potentially useful scientific information).

evidence. Instead, the Court created a new gatekeeping role for the judge[35] and a two-step test designed to govern the admissibility inquiry.[36] As a first step, judges must determine whether evidence is "scientific knowledge."[37] This "requires judges to critique scientific evidence and separate the wheat of valid scientific methodology from the chaff of chicanery."[38] The judge's focus, according to the *Daubert* Court, must be "solely on principles and methodology, not on the conclusions that they generate."[39] To assist judges who would now need to locate proffered evidence on the continuum between reliable and unreliable, Justice

---

[35] Justice Blackmun was joined by Justices Kennedy, O'Connor, Scalia, Souter, Thomas, and White in his creation of the judicial gatekeeper, while Chief Justice Rehnquist and Justice Stevens dissented. It should be noted that the immediate response among the federal judiciary to *Daubert* was generally negative. In the words of one commentator, "Many federal judges believe *Daubert* has made their lives more difficult. . . . They are going to have to give a more reasoned statement about why they are letting in evidence. . . . They can't do it on a rubber-stamp basis the way some of them did it in the past." Rorie Sherman, *Judges Learning* Daubert*: 'Junk Science' Rule Used Broadly*, NAT'L L.J., Oct. 4, 1993, at 3 (quotation marks omitted) (describing judicial discomfort with *Daubert* and quoting U.S. District Judge Jack B. Weinstein as saying, "After all, . . . we're not scientists").

[36] *See Daubert*, 509 U.S. at 589 (explaining that the trial judge must ensure that all scientific testimony or admitted evidence is not only relevant but reliable). The Sixth Circuit has interpreted the *Daubert* decision as follows:

> *Daubert* thus requires trial courts to perform a two-step inquiry. First, the court must determine whether the expert's testimony reflects scientific knowledge, that is, the court must 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. Second, the court must ensure that the proposed expert testimony is relevant to the task at hand and will serve to aid the trier of fact.

United States v. Smithers, 212 F.3d 306, 313 (6th Cir. 2000) (citations and quotation marks omitted) .

[37] According to Justice Blackmun, "[f]aced with a proffer of expert scientific testimony, then, the trial judge must determine at the outset . . . whether the expert is proposing to testify to . . . scientific knowledge." *Daubert*, 509 U.S. at 592. However, "scientific knowledge" is defined only vaguely by the *Daubert* Court as "an inference or assertion . . . derived by the scientific method." *Id.* at 590. Despite the fact that "scientific knowledge" and "scientific method" are critical components of the *Daubert* decision, the Court does little to elucidate these terms of art or to distinguish one from the other. In fact, "scientific method" is mentioned only twice by the Court. Justice Blackmun defines "scientific method" as scientific knowledge that "implies a grounding in the methods and procedures of science." *Id.* at 590. Then he describes scientific method as "based on generating hypotheses and testing them to see if they can be falsified." *Id.* at 593.

[38] Erica Beecher-Monas, *Blinded by Science: How Judges Avoid the Science in Scientific Evidence*, 71 TEMP. L. REV. 55, 62 (1998).

[39] *Daubert*, 509 U.S. at 595. Four years after *Daubert*, the Court acknowledged that the task of separating principles, methodology, and conclusions is not as simple as they had previously assumed. *See* United States v. Joiner, 522 U.S. 136, 146 (1997) (concluding that in science "conclusions and methodology are not entirely distinct from one another").

Blackmun outlined four criteria: (1) testability; (2) peer review and publication; (3) error rate; and (4) general acceptance.[40] The second step mandated that judges decide whether the evidence "fits" or is relevant to the facts at issue.[41]

Over the past decade, *Daubert* has transformed judicial decision making on questions of science and law in the federal courts and the thirty states that have adopted *Daubert* in whole or in part.[42] Even in states that retain a *Frye*-type

---

[40] *See Daubert*, 509 U.S. at 594.

[41] The *Daubert* Court identified the second step of the inquiry as determining whether the "scientific knowledge . . . will assist the trier of fact to understand or determine a fact in issue." *Id.* at 592. This is essentially identical to the requirements under Federal Rule of Evidence 702 that an expert testifying to "scientific . . . knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue." FED. R. EVID. 702.

[42] *See* Alabama—So. Energy Homes, Inc. v. Washington, 774 So. 2d 505, 516–17 (Ala. 2000) (acknowledging that the legislature has used *Daubert* with respect to DNA evidence, but not explicitly switching the standard from *Frye* to *Daubert* for other evidence); Alaska—State v. Coon, 974 P.2d 386, 388–99 (Alaska 1999) (adopting the *Daubert* standard); Arizona—Logerquist v. Mcvey, 1 P.3d 113, 132 (Ariz. 2000) (retaining the *Frye* standard); Arkansas—Moore v. State, 915 S.W.2d 284, 293–94 (Ark. 1996) (recognizing the *Daubert* standard but not expressly adopting it); California—People v. Leahy, 882 P.2d 321, 323–24 (Cal. 1994) (refusing to adopt *Daubert* and noting that California has long held to the *Frye* standard and would continue to do so); Colorado—People v. Shreck, 22 P.3d 68 (Colo. 2001) (adopting a three-part standard based on reliability, qualifications, and usefulness); Connecticut—State v. Porter, 698 A.2d 739, 751 (Conn. 1997) (adopting the *Daubert* standard); Delaware—Bell Sports, Inc. v. Yarusso, 759 A.2d 582, 588–90 (Del. 2000) (expressly adopting *Daubert*); Florida—Brim v. State, 695 So. 2d 268, 271–72 (Fla. 1997) (rejecting *Daubert*); Georgia—Jordan v. Ga. Power Co., 466 S.E.2d 601, 604–05 (Ga. 1995) (applying state law and not adopting *Daubert*); Hawaii—State v. Fukusaku, 946 P.2d 32, 42 (Haw. 1997) (refusing to follow *Daubert*); Idaho—State v. Trevino, 980 P.2d 552, 557–58 (Idaho 1999) (adopting the *Daubert* standard); Illinois—Donaldson v. Cent. Ill. Pub. Serv. Co., 767 N.E.2d 314, 323 (Ill. 2002) (reaffirming that Illinois follows the *Frye* standard); Indiana—Sears Roebuck & Co. v. Manuilov, 742 N.E.2d 453, 462 (Ind. 2001) (retaining the *Frye* standard); Iowa—Leaf v. Goodyear Tire & Rubber Co., 590 N.W.2d 525, 530–33 (Iowa 1999) (adopting a limited application of *Daubert*); Kansas—State v. Canaan, 964 P.2d 681, 691–92, 694 (Kan. 1998) (retaining the *Frye* standard); Kentucky—Mitchell v. Commonwealth of Ky., 908 S.W.2d 100 (Ky. 1995) (adopting *Daubert* expressly); Louisiana—State v. Ledet, 792 So. 2d 160 (La. 2001) (adopting the *Daubert* standard); Maine—State v. McDonald, 718 A.2d 195 (Me. 1998) (adopting *Daubert*); Maryland—Hutton v. State, 663 A.2d 1289, 1295–96 n.10 (Md. 1995) (determining that Maryland will still follow the *Frye* standard despite the fact that Maryland's Rules of Evidence are patterned after the Federal Rules of Evidence and were passed into legislation after the *Daubert* decision); Massachusetts—Commonwealth v. Senior, 744 N.E.2d 614 (Mass. 2001) (applying various *Daubert* factors); Minnesota—State v. Klawitter, 518 N.W.2d 577, 585 n.3 (Minn. 1994) (noting that the *Frye* standard has been utilized before and after *Daubert* although expressing that "we do not address the effect of the *Daubert* decision on the use or application of the *Frye* rule in Minnesota"); Mississippi—Gleeton v. State, 716 So. 2d 1083, 1087 (Miss. 1998) (retaining the *Frye* standard); Missouri—Callahan v. Cardinal Glennon Hosp., 863 S.W.2d 852 (Mo. 1993) (continuing to apply *Frye*); Montana—State v. Moore, 885 P.2d 457 (Mont. 1994) (adopting the *Daubert* standard); Nebraska—Sheridan v.

"general acceptance" admissibility standard, many state court judges report that *Daubert* has had a powerful influence on their decisions.[43] In state courts, judges have held pretrial hearings or developed other methods for determining the relevance and, more significantly, the validity of proffered scientific evidence.[44]

Catering Mgmt., Inc., 566 N.W.2d 110, 113 (Neb. 1997) (retaining the *Frye* standard); Nevada—Dow Chem. Co. v. Mahlum, Inc. 973 P.2d 842 (Nev. 1999) (applying various *Daubert* factors); New Hampshire—State v. Cort, 766 A.2d 260 (N.H. 2000) (applying various *Daubert* factors); New Jersey—State v. Harvey, 699 A.2d 596, 621 (N.J. 1997) (applying various *Daubert* factors); New Mexico—State v. Anderson, 881 P.2d 29 (N.M. 1994) (adopting the *Daubert* standard); New York—People v. Wernick, 674 N.E.2d 322, 324 (N.Y. 1996) (retaining the *Frye* standard); North Carolina—State v. Goode, 461 S.E.2d 631, 639, 641 (N.C. 1995) (adopting the *Daubert* standard); North Dakota—City of Fargo v. McLaughlin, 512 N.W.2d 700, 705 n.2 (N.D. 1994) (retaining the *Frye* standard); Ohio—Miller v. Bike Athletic Co., 687 N.E.2d 735 (Ohio 1998) (adopting the *Daubert* standard); Oklahoma—Torres v. State, 962 P.2d 3, 22 (Okla. Crim. App. 1998) (holding that *Daubert* is "not applicable to non-scientific evidence"); Taylor v. State, 889 P.2d 319, 328 (Okla. Crim. App. 1995) (adopting, in this criminal case, the *Daubert* standard as it applies to novel or new "scientific or technical evidence"); Oregon—State v. Brown, 687 P.2d 751 (Or. 1984) (adopting its own standard to determine whether scientific evidence is probative); State v. O'Key, 899 P.2d 663, 680 (Or. 1995) (retaining the *Brown* standard but noting that trial courts "should . . . find *Daubert* instructive"); Pennsylvania—Commonwealth v. Arroyo, 723 A.2d 162, 170 n.10 (Pa. 1999) (retaining the *Frye* standard); Rhode Island—State v. Quattrocchi, 681 A.2d 879, 884 n.2 (R.I. 1996) (adopting the *Daubert* standard); South Carolina—State v. Council, 515 S.E.2d 508, 518 (S.C. 1999) (using factors similar to, but not specifically adopting, the *Daubert* factors); South Dakota—State v. Hofer, 512 N.W.2d 482 (S.D. 1994) (adopting the *Daubert* standard); Tennessee—McDaniel v. CSX Transp., Inc., 955 S.W.2d 257 (Tenn. 1997) (adopting the *Daubert* standard); Texas—E.I. du Pont Nemours & Co. v. Robinson, 923 S.W.2d 549 (Tex. 1995) (adopting the *Daubert* standard); Utah—State v. Butterfield, 27 P.3d 1133 (Utah 2001) (holding that the test for admissibility requires threshold showing of "inherent reliability"); Vermont—State v. Brooks, 643 A.2d 226 (Vt. 1993) (adopting the *Daubert* decision); Virginia—Spencer v. Commonwealth, 393 S.E.2d 609, 621 (Va. 1990) (declining expressly to follow *Frye*, but not adopting *Daubert*); Washington—State v. Copeland, 922 P.2d 1304, 1310 (Wash. 1996) (retaining the *Frye* standard); West Virginia—Wilt v. Buracker, 443 S.E.2d 196 (W. Va. 1993) (adopting the *Daubert* decision); Wisconsin—State v. Peters, 534 N.W.2d 867 (Wis. Ct. App. 995) (basing admissibility on a three-part relevance test); Wyoming—Bunting v. Jamieson, 984 P.2d 467 (Wyo. 1999) (adopting the *Daubert* standard). Although the Michigan Supreme Court has not addressed the issue, the consensus among the lower courts favors *Frye*. *See, e.g.*, Nelson v. Am. Sterilizer Co., 566 N.W.2d 671, 673–74 (Mich. Ct. App. 1997); The District of Columbia has not yet adopted Federal Rule of Evidence 702, and there has been no majority opinion that has addressed *Daubert*. *Cf.* Taylor v. United States, 661 A.2d 636, 651–52 (D.C. 1995) (Newman, S.J., dissenting) (urging the adoption of Federal Rule 702 and *Daubert*).

[43] *See* Gatowski et al., *supra* note 7, at 443 (describing how 94% of state court judges surveyed find *Daubert* has either "some value" or "a great deal of value" for their decision-making process on questions involving scientific evidence regardless of whether *Daubert* or *Frye* governs admissibility in their jurisdiction).

[44] Although the *Daubert* Court used the word "reliable" to refer to the quality of the scientific evidence, I have argued elsewhere that this reflects a misunderstanding of this

In states that require *Daubert* hearings, the burden is on proponents to establish by a preponderance of the evidence that the admissibility requirements have been met.[45] Proponents, however, need not show that their experts' conclusions are *correct*.[46] Courts need only be persuaded that the science supporting the conclusions are sufficiently valid and relevant.

B. *Recent Empirical Research Reveals that Many Judges Cannot Assess Scientific Validity*

Chief Justice Rehnquist demonstrated remarkable candor and prescience when he wrote the following for the dissent in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*:[47] "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 suppose some of them will be too."[48] A very recent survey of four hundred state judges demonstrates that the vast majority do not understand even the most basic scientific concepts described by the *Daubert* Court.[49]

This new national survey is the most comprehensive effort to collect empirical data assessing how well judges understand and apply *Daubert*.[50] The primary purpose of the study was to:

---

scientific term of art. *See* Moreno, *supra* note 27, at 1065–70 (describing how "reliability" refers only to the reproducibility of data, even if the data is wrong, while "validity" connotes a connection between the theory or conclusions and the empirical world).

[45] *See* Bourjaily v. United States, 483 U.S. 171, 172–73 (1987).

[46] According to the Eleventh Circuit, "the proponent of the testimony does not have the burden of proving that it is scientifically correct, but that by a preponderance of the evidence, it is reliable." Allison v. McGhan Med. Corp., 184 F.3d 1300, 1312 (11th Cir. 1999). However, the idea of a correct, or completely valid, scientific conclusion is itself misleading. Scientific theories are almost never categorized as valid or invalid because they are rarely wholly accurate or wholly inaccurate explanations of the empirical world. Thus, we should also recognize that " '[v]alidity' in science is not a binary attribute, like pregnancy." FOSTER & HUBER, *supra* note 1, at 17 (discussing issues of scientific uncertainty and the limited ability of scientists to speak in terms of absolutes). Scientific validity is better understood as a matter of degree rather than in absolute terms. If the court, therefore, determines that the scientific reliability of a theory is low, then its validity is suspect. A high level of reliability, on the other hand, does not establish the validity of a particular scientific theory or test.

[47] 509 U.S. 579 (1993).

[48] *Id.* at 600 (Rehnquist, C.J., dissenting).

[49] This national survey involved judges throughout the country in states that have adopted *Daubert*, states that have adopted a modified version of *Daubert*, and states that continue to apply *Frye*. *See* Gatowski et al., *supra* note 7, at 439.

[50] The study involved 400 state court judges. Part I of the study (a structured telephone interview) obtained a 71% response rate. Part II involved either a structured telephone interview or a follow-up written questionnaire and obtained an 81% response rate. *See id.* at 433–35.

> assess the level to which the judiciary understand the scientific meaning of the *Daubert* guidelines and how they might apply them when evaluating the admissibility of scientific evidence. In addition to assessing the scientific literacy of judges, the survey also asked respondents for their opinions about the relevance and utility of the *Daubert* criteria to the judicial gate keeping role and admissibility decision-making process.[51]

With respect to the first goal, the researchers concluded that *Daubert* is neither accurately nor consistently applied in the state courts.[52] With respect to the second goal, they found that despite obvious confusion about how to apply *Daubert*, the vast majority of state court judges (94%) in both *Frye* and *Daubert* jurisdictions report that they find *Daubert* valuable to their decision-making process, with 55% reporting that *Daubert* provides a " 'great deal' of value."[53]

This study provides information that has never previously been collected, analyzed, or published.[54] Previously, the limited amount of empirical research performed in this area involved retrospective analyses of published judicial opinions.[55] These earlier studies inferred conclusions about the utility and relevance of *Daubert* to legal decision-making.[56] Researchers involved in the current study highlighted the deficiencies inherent in earlier research methodologies:

> While providing important insight regarding the influence of *Daubert*, an empirical analysis of published case law is, by its very nature, restricted to an analysis of post hoc justifications of those writing a decision in a particular case and does not fully capture the judicial decision-making process. Although an empirical analysis of case law provides important data about judges' normative, case specific reasoning, research has demonstrated that there may be significant differences between published and unpublished cases, and that these differences may be dependent upon the case characteristics analyzed and the legal questions involved.[57]

Direct questioning of judges offers numerous advantages over previous methodologies. First, it eliminates the need for, and inherent unreliability of, inferred judicial motives.[58] Second, it can reveal more directly judges' thoughts

---

[51] *Id.* at 438.

[52] *See id.* at 443.

[53] *Id.*

[54] *See id.* at 433–35.

[55] *See* Gatowski et al., *supra* note 7, at 434–35.

[56] *See id.*

[57] *Id.* (citation omitted).

[58] *See id.* at 435.

and level of scientific comprehension.[59] Third, the question and answer process used by the researchers enabled judges to describe how they use the *Daubert* criteria to assess scientific evidence, unencumbered by their assessment of the facts or law in a particular case.

The results of the new study are dramatic. Researchers found that the overwhelming majority of judges have no working understanding of two of the four *Daubert* criteria.[60] For example, while 88% of the judges reported that "falsifiability" is a useful guideline for determining the quality of proffered scientific evidence, 96% of these same judges lacked even a basic understanding of this core scientific concept.[61] Nine years ago, the *Daubert* Court concluded that falsifiability was "a key question to be answered in determining whether a theory or technique is scientific knowledge that will assist the trier of fact"[62] and defined falsifiability as "whether [the theory] can be (and has been) tested."[63] In a law review article by Professor Mark Green, quoted by the Supreme Court in *Daubert*,[64] falsifiability is further defined as the theory that "knowledge is gained by attempting to disprove or falsify a hypothesis based on empirical investigation. Scientific methodology today is based on generating hypotheses and testing them to see if they can be falsified; indeed, this methodology is what distinguishes science from other fields of human inquiry."[65] Surveyed judges were not expected to demonstrate even this level of comprehension. In fact, responses as simple as "I would want to know to what extent the theory has been properly and sufficiently tested and whether or not there has been research that has attempted to prove the theory to be wrong" or "if it is not possible to test the evidence then it would weigh heavily with me in my decision" were deemed accurate.[66] Only 14 judges out of 352, demonstrated even this level of understanding.[67]

---

[59] *See id.*

[60] *See id.* at 444–46.

[61] *See* Gatowski et al., *supra* note 7, 444–45.

[62] Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 593 (1993).

[63] *Id.*

[64] *See Daubert*, 509 U.S. at 586 n.4 (citing Michael D. Green, *Expert Witnesses and Sufficiency of Evidence in Toxic Substances Litigation: The Legacy of Agent Orange and Bendectin Litigation*, 86 NW. U. L. REV. 643 (1992)).

[65] Green, *supra* note 64, at 645; *see also* KENNETH R. FOSTER & PETER W. HUBER, JUDGING SCIENCE: SCIENTIFIC KNOWLEDGE AND THE FEDERAL COURTS 38 (1999) (noting that under the "view of science [adopted by the *Daubert* Court] hypotheses are never affirmatively proved, they are only falsified . . . [b]ut a hypothesis that repeatedly withstands attempts to falsify it will become accepted by the scientific community").

[66] Gatowski et al., *supra* note 7, at 444.

[67] *Id.* at 444–45. In a discussion of why judges must understand falsifiability, Professor Faigman notes that:

> judges must develop sufficient scientific literacy to recognize research designed to truly test a hypothesis as compared to research designed merely to supply impressive looking

Similarly, 91% of the judges reported that they found error rates useful for determining the quality of proffered scientific evidence.[68] Here again, judges do not seem to understand the scientific concepts they routinely employ. The *Daubert* Court cautioned that "in the case of a particular scientific technique, the court ordinarily should consider the known or potential rate of error and the existence and maintenance of standards controlling the technique's operation."[69] However, judges have misunderstood the definition of error rates and, therefore, their significance. When error rates are used to assess the validity of a scientific methodology, they can include false negative errors (when an experimenter misses a real effect), false positive errors (when an experimenter perceives an effect that did not occur), and sampling errors (when an experimenter extrapolates from a small sample to a large population).[70] Only 4% of the judges who reported that error rates were useful, demonstrated a fundamentally accurate understanding of the definition of error rates. As with falsifiability, researchers did not expect a highly sophisticated level of comprehension. Responses defined as accurate included: "it would seem that if a theory or procedure has too high an error rate it would have to be rejected because the risk is too high of being wrong" and "I would want to know about the probability of making a mistake."[71] Only 15 judges out of 364, had even this level of comprehension.

Finally, the study found that judges scored much higher in their basic comprehension of the last two *Daubert* criteria's definitions: peer review and publication (71%) and general acceptance (82%).[72] The researchers concluded that "[t]he survey findings strongly suggest that judges have difficulty operationalizing the *Daubert* criteria and applying them, especially with respect to falsifiability and error rate."[73] The researchers also noted that despite specific efforts by interviewers aimed at allowing judges to express their level of comprehension, "it seems likely that the ambiguity of the [judges'] responses may reflect a genuine lack of understanding of these scientific concepts."[74]

The real world implications of this study are profound. As the world grows more scientifically complex, the fact that many judges seem to lack even basic

---

> graphs and imposing numbers to a researcher's theory. In other words, judges (and lawyers) must be able to distinguish the methods of science from those methods that merely imitate science.

FAIGMAN ET AL., SCIENCE IN THE LAW: STANDARDS, STATISTICS AND RESEARCH ISSUES 28 (2002) (footnotes omitted).

[68] Gatowski et al., *supra* note 7, at 445.

[69] *Daubert*, 509 U.S. at 594 (citations omitted).

[70] *See* FOSTER & HUBER, *supra* note 1, at 75–76 (defining false positive and false negative errors and describing how they can result in sampling error).

[71] *See* Gatowski et al., *supra* note 7, at 445–47.

[72] *See id.* at 447–48.

[73] *Id.* at 452.

[74] *Id.*

familiarity with the scientific process raises serious concerns. A decade after *Daubert*, state courts have demonstrable, ongoing, and systemic problems assessing the quality of scientific evidence. It appears that the recent proliferation of academic literature on science and law has done little to educate judges who must make difficult decisions about the admissibility of scientific evidence in real cases. If judges do not understand the *Daubert* criteria, they cannot hope to make meaningful, accurate, or consistent assessments of scientific evidence. These problems will likely be exacerbated in the future following the Supreme Court's recent expansion of *Daubert* to all technical and specialized knowledge through *Kumho Tire Co. v. Carmichael*.[75]

III. THE CURRENT LEGAL STANDARD GOVERNING THE ADMISSION OF SCIENTIFIC EVIDENCE—THE POST-*DAUBERT* CASES

Daubert and its progeny are based on the assumption that jurors must be shielded from scientific-sounding evidence that is either irrelevant or invalid science. One commentator has noted that:

> *Daubert*'s underlying rationale is . . . sound . . . : lay jurors should not be exposed to unfiltered scientific or technical testimony that may adversely influence their findings of fact. But this rationale is built on two underlying assumptions: (1) that the trial judge is more knowledgeable in assessing complex scientific testimony than is the . . . lay juror, and (2) that each judge brings to the specific task of gatekeeping a general attitude or philosophy concerning the level of scrutiny appropriate for scientific gatekeepers.[76]

The two most important post-*Daubert* cases from the Supreme Court reinforce the central role of the trial judge by insulating most admissibility decisions from appellate review and vastly expanding the application of *Daubert* to non-scientific expertise.

---

[75] 526 U.S. 137, 141 (1999) (holding that the *Daubert* gate-keeping role "applies not only to [expert] testimony based on 'scientific' knowledge, but also to testimony based on 'technical' and 'other specialized' knowledge").

[76] Walsh, *supra* note 4, at 143.

In the first case, *Joiner v. General Electric Co.*,[77] Chief Justice Rehnquist, writing for the majority, held that abuse of discretion is the appropriate standard of review for all evidentiary rulings, including the exclusion of scientific expert testimony.[78] The *Joiner* Court expanded our evolving understanding of the proper admissibility standard, noting that scientific "conclusions and methodology are not entirely distinct from one another."[79] Chief Justice Rehnquist specifically cautioned judges attempting to apply *Daubert* that "nothing in either *Daubert* or the Federal Rules of Evidence requires a district court to admit opinion evidence which is connected to existing data only by the *ipse dixit* of the expert."[80] Judges applying *Daubert* after *Joiner*, therefore, must assess the scope of the "analytical gap between the data and the opinion proffered" to determine if there is a sufficiently close correlation for the evidence to be admitted.[81]

The most significant clarification of *Daubert* by the Supreme Court occurred two years after *Joiner*, in *Kumho Tire Co. v. Carmichael*.[82] *Kumho* restated the *Daubert* admissibility standard but made two explicit additions: (1) *Kumho* expanded the *Daubert* gatekeeping role to include testimony by experts with scientific, technical, or other specialized non-empirical knowledge[83] and (2) *Kumho* added a requirement that experts employ "the same level of intellectual rigor" in the courtroom as in their fields of research.[84] However, Justice Breyer's

---

[77] 522 U.S. 136 (1997). In 1997, the Supreme Court granted certiorari in *General Electric Co. v. Joiner*, 522 U.S. 136 (1997), to resolve the question left open by *Daubert* of the appropriate standard for appellate review of a trial court's decision to admit or exclude scientific evidence. In *Joiner*, the plaintiff claimed that exposure to polychlorinated biphenyls ("PCBs") had caused his lung cancer. *Joiner v. General Electric Co.*, 864 F. Supp. 1310, 1314 (N.D. Ga. 1994). To support his claim, the plaintiff offered four epidemiological studies that purportedly established a causal link between defendant's PCBs and plaintiff's cancer. *Joiner*, 522 U.S. at 145. The district court reviewed the plaintiff's four studies and found that: (1) the first study did not conclude that PCBs had caused lung cancer among the workers they examined; (2) the second study found that there was a slightly increased incidence of lung cancer among workers at a PCB plant, but that the increase was not statistically significant; (3) the third study did not mention PCBs; and (4) the fourth study's subjects had been exposed to numerous potential carcinogens. *Id.* at 145–46. After excluding all of the plaintiff's scientific expert testimony, the district court granted summary judgment for the defendant. The Eleventh Circuit used a "stringent standard of review" to reverse the district court. *Id.* at 141–43.

[78] *See Joiner*, 522 U.S. at 143.

[79] *Id.* at 146.

[80] *Id.*

[81] *See id.* at 146.

[82] 526 U.S. 137 (1999).

[83] *See id.* at 148; *see also* Edward J. Imwinkelried, *The Taxonomy of Testimony Post-*Kumho*: Refocusing on the Bottomlines of Reliability and Necessity*, 30 CUMB. L. REV. 185, 209 (2000) (noting that, prior to *Kumho Tire*, "[t]he objective validity of a non-scientific expert's premises was essentially exempt from any scrutiny").

[84] *See Kumho Tire*, 526 U.S. at 152.

majority opinion in *Kumho* did more. *Kumho* was the first time the Court acknowledged and addressed the problems that judges have had understanding and implementing *Daubert*.

Judges' problems with *Daubert* may be attributable in part to the text of the case itself. In the view of one commentator:

> The problem is that *Daubert* describes evidentiary reliability rather than defines it. Moreover, *Daubert*'s description is imprecise, couched in scientific jargon, and, accordingly, difficult to apply even when dealing with testimony that is unquestionably scientific. The *Daubert* majority leads the reader through a legal and scientific maze, entering at rule 702 and its "scientific . . . knowledge" requirement, and exiting, after many twists and turns, at "evidentiary reliability" and factors that are indicators of reliability. Like a maze, the twists and turns add confusion, not clarity.[85]

I have argued elsewhere that *Kumho* reflects the Court's effort to clarify *Daubert* by correcting two inherent structural problems.[86] The first problem is one of interpretation. The primacy of the general reliability/validity step of the *Daubert* test seems to require judges to first determine whether proposed expert testimony is "scientific knowledge," *before* exploring its relevance. This structure appears to distort the admissibility decision by demanding that judges assess a potentially infinite amount of scientific evidence, most of which is not relevant to the instant facts. The second problem is one of application: Judges operating the *Daubert* standard may mistakenly assume that, because they have little expertise evaluating competing scientific theories, they should admit all but the most patently bogus scientific evidence and allow the jurors to resolve discrepancies as questions of weight.[87]

---

[85] Robert J. Goodwin, *The Hidden Significance of* Kumho Tire Co. v. Carmichael*: A Compass for Problems of Definition and Procedure Created by* Daubert v. Merrell Dow Pharmaceuticals, Inc., 52 BAYLOR L. REV. 603, 613 (2000).

[86] *See* Moreno, *supra* note 27, at 1052–55.

[87] This concern is addressed in the Advisory Committee Notes to the May 2000 amendments of Federal Rule of Evidence 702, which state that "[a] review of the case law after *Daubert* shows that the rejection of expert testimony is the exception rather than the rule." FED. R. EVID. 702 advisory committee's note; *see also* Beecher-Monas, *supra* note 38, at 58 ("All too often, however, courts continue to evade the science issues. In far too many jurisdictions, judges are turning a blind eye to the science involved in the evidence before them."); David L. Faigman et al., *How Good Is Good Enough?: Expert Evidence Under* Daubert *and* Kumho, 50 CASE W. RES. L. REV. 645, 665 (2000) ("In the forensic context, courts have long admitted a surfeit of expertise with little or no evaluation of the foundation upon which the opinion rests."); Jay P. Kesan, *Drug Development: Who Knows Where the Time Goes?: A Critical Examination of the Post-*Daubert *Scientific Evidence Landscape*, 52 FOOD & DRUG L.J. 225, 239–40 (1997) (reviewing numerous post-*Daubert* cases and concluding that "the quantum of scientific information that must undergird an expert's methodology to render it scientifically valid and admissible under *Daubert* is quite minimal").

Justice Breyer resolves both problems by modeling the appropriate judicial inquiry, so that it is almost exclusively focused on the fit/relevance prong of *Daubert*.

> [T]he specific issue before the [district] court was not the reasonableness *in general* of a tire expert's use of a visual and tactile inspection . . . [but was instead] the reasonableness of using such an approach . . . to draw a conclusion regarding *the particular matter to which the expert testimony was directly relevant*.[88]

Kumho narrows the scope of *Daubert* and reflects a deliberate effort by the Court to shift its focus of the judicial inquiry away from general scientific validity and towards an evaluation of the specific scientific evidence, inferences, and conclusions drawn from this evidence to determine whether they are relevant to the dispute.[89] This reading is supported by the *Kumho* Court's articulation of the proper standard. "[T]he question before the trial court was specific, not general. The trial court had to decide whether *this particular expert* had sufficient specialized knowledge to assist jurors in deciding *the particular issues* in the case."[90] The cases that follow Kumho are consistent with this interpretation.[91]

---

[88] *Kumho Tire*, 526 U.S. at 153–54.

[89] The very few other legal scholars who have noted the *Kumho Tire* Court's almost exclusive focus on the relevance inquiry have occasionally referred to this as emphasis on the "task at hand." *See, e.g.*, D. Michael Risinger, *Defining the "Task at Hand": Non-Science Forensic Science After* Kumho Tire Co. v. Carmichael, 57 WASH. & LEE L. REV. 767 (2000). Professor Risinger observes:

> what is clearly not consistent with *Kumho Tire* is any attempt to approach an issue of reliability globally. That is, reliability cannot be judged globally, "as drafted," but only specifically, "as applied." The emphasis on the judgment of reliability as it applies to the individual case, to the "task at hand," runs through the opinion like a river.

Id. at 773.

[90] *Kumho Tire*, 526 U.S. at 156 (emphasis added) (citations omitted).

[91] *See, e.g.*, United States v. Brumley, 217 F.3d 905, 911 (7th Cir. 2000) (finding that "[t]he Supreme Court in *Kumho Tire* explained that the *Daubert* 'gatekeeper' factors had to be *adjusted to fit the facts of the particular case at issue*, with the goal of testing the reliability of the expert opinion") (emphasis added); Seatrax, Inc. v. Sonbeck Int'l., Inc., 200 F.3d 358, 372 (5th Cir. 2000) ("[W]hether Daubert's suggested indicia of reliability apply to any given testimony depends on the nature of the issue at hand, the witness's particular expertise, and the subject of the testimony. It is a fact-specific inquiry.") (citation omitted); United States v. Smithers, 212 F.3d 306, 315 (6th Cir. 2000) (noting that the *Kumho Tire* court engaged in a thorough reexamination of the technology relevant to the facts that had been presented to the district court); United States v. Horn, 185 F. Supp. 2d 530, 554 (D. Md. 2002) (stating that "judges do not determine the reliability of scientific or technical issues in the abstract but rather in the context of deciding a specific dispute").

IV. PRACTICAL PROBLEMS DEMAND PRACTICAL SOLUTIONS

A. *The Importance of Understanding Science*

Judges do not need to become trained scientists to achieve accurate and consistent legal decision-making in cases involving scientific evidence. They need to become savvy consumers of the scientific evidence that comes before them. This process begins when judges identify scientific ideas they do not understand[92] and focus on assessing evidence, conclusions, and inferences drawn from this evidence that inform their understanding of the dispute.[93]

Judges, who seek guidance from the relevant academic literature may be ill-served by scholarly articles that tend to treat all science as a single discipline distinguished only by its classification as valid or junk.

> The rejection of a simple dichotomy between "good" and "bad" science facilitates discussion in a number of areas otherwise precluded. For instance, questions relating to the efficacy of various sciences, their objectives, and the ethics of their practitioners can be examined in more specific local terms, freed from the need to anchor them to over-arching, unworkable, mythological images of science.[94]

Moreover, recent empirical evidence demonstrates that non-scientists have more difficulty understanding and employing methodological reasoning if it is taken out of context. One study has shown that non-scientists can correctly use certain rules, such as conditional probability, only when they are in context, that is, when subjects are faced with a concrete task.[95] However, when these same subjects are "faced with an abstract problem that has the same logical structure . . . their performance is very poor."[96] If a judge is

> incapable of understanding problems such as statistical representativeness, confounded variables, and conditional probabilities, then he or she will not be able to grasp the reasoning behind an expert opinion, even if it is clearly

---

[92] *See supra* Part II (discussing the empirical evidence identifying the *Daubert* criteria that cause the most problems for judges).

[93] *See supra* Part III (describing how *Kumho Tire* narrowed the scope of the *Daubert* admissibility inquiry).

[94] Edmond & Mercer, *supra* note 14, at 33.

[95] Neil Vidmar & Shari Seidman Diamond, *Juries and Expert Evidence: From the Nineteenth to the Twenty-First Century*, 66 BROOK. L. REV. 1121, 1136 (2001) (referencing the work of social psychologist Harold Kelley).

[96] *Id.*

> explained and examined during direct and cross-examination of the expert witness.[97]

This suggests that judges might learn by exploring a real scientific controversy. Child abuse cases provide context for an analysis of the type of complex medical expert testimony that must frequently be evaluated by courts.

B. *The Three Stages of a Child Abuse Case*

1. *Stage One: Detecting and Diagnosing the Physical Abuse of Children*

> In physical [child] abuse cases, the victim's injured body often provides the most compelling evidence . . . .
>
> Professor John E.B. Meyers[98]

---

[97] *Id.* at 1135.

[98] John E.B. Meyers, *Child Abuse: Introduction*, 28 PAC. L.J. 1, 1 (1996).

Child abuse is defined by state law.[99] Thus, the statutory definition of abuse in a particular jurisdiction defines the crime, which in turn determines which

---

[99] ALA. CODE §§ 26-14-1(1) to -1(3), -7.2(a) (Supp. 2002); ALASKA STAT. §§ 47.17.020(d), .290 (Michie 2002); ARIZ. REV. STAT. ANN. § 8-201(l), (2), (6), (8), (12), (13)(a)(i)–(ii), 13(b), (21) (West Supp. 2002); ARK. CODE ANN. § 12-12-503, (Michie Supp. 2001); CAL. PENAL CODE §§ 11165.1–.6, .12 (West 2000 & Supp. 2003); COLO. REV. STAT. ANN. § 19-1-103(l)(a)–(b), (18), (27), (32), (35), (37), (66), (67), (82), (94), (97), (101), (104), (108), (111) (West 2002); §§ 19-3-102, -103; CONN. GEN. STAT. ANN. § 46b-120(l), (2), (4), (9), (10) (West Supp. 2002); § 17a-104 (West 1998); DEL. CODE ANN. tit. 16, §§ 902, 913 (Supp. 2002); D.C. CODE ANN. § 4-1301.02(1), (4)–(6), (8), (16)–(18), (20) (Supp. 2002); § 16-2301(9), (23)–(25) (2001); FLA. STAT. ANN. § 39.01(l), (2), (10), (12), (14), (19), (27), (30), (31), (43), (45), (47), (52), (63), (71) (West Supp. 2003); GA. CODE ANN. § 19-7-5(b), (1999); HAW. REV. STAT. § 350-1 (1998); IDAHO CODE § 16-1602 (Michie 2001); 325 ILL. COMP. STAT. ANN. 5/3 (West Supp. 2002); IND. CODE ANN. §§ 31-9-2-0.5, -123, -132 (West 1999 & Supp. 2002); §§ 31-34-1-1 to -6, -8 to -12, -14, -15 (West 1999); § 35-46-1-3 (West 1998); IOWA CODE ANN. § 232.68(2), (4), (5), (7) (West Supp. 2002); KAN. STAT. ANN. § 38-1502(a)–(h), (i)–(r), (t), (v)–(z), (cc) (2002); §§ 21-3501(l), (2), (4), -3502(a), -3503(a), -3504(a), -3510(a), -3511, -3516(a), (b)(1), -3602, -3603 (Supp. 2001); § 38-1502(cc)(3); KY. REV. STAT. ANN. § 600.020(1), (2), (6), (8), (18), (24), (37), (38), (42), (44), (54)–(56) (Michie Supp. 2002); LA. CHILDREN'S CODE ANN. art. 603(1), (3), (5), (7), (7.1), (9), (10), (14), (17) (West Supp. 2003); ME. REV. STAT. ANN. tit. 22, § 4002(l), (1-A), (1-B), (6), (9), (9-B), (9-C), (10), (11) (West 1992 & Supp. 2002); § 4010(1); MD. CODE ANN. FAM. LAW § 5-701 (1991); MASS. GEN. LAWS ANN. ch. 119, § 21 (West 2002); § 51A (West Supp. 2002); ch. 209, § 38 (West 2002); MICH. COMP. LAWS ANN. § 722.622(b)–(f), (i)–(k), (o), (p), (r), (s), (v), (w) (West 2002); § 722.628(3)(c); §722.634; MINN. STAT. ANN. § 260C.007 subd. 3, 4, 8, 12, 14, 15, 18, 21–26 (West Supp. 2003); § 626.556 subd. 2(a)–(c), (d), (e), (k), (l), (m), Subd. 11d(a); MISS. CODE ANN. § 43-21-105(d) to (g), (i) to (n), (v) (Supp. 2002); MO. ANN. STAT. §§ 210.110, .115(3) (West Supp. 2003); MONT. CODE ANN. § 41-3-102(1) to (4), (5) to (9), (11) to (15), (17) to (18), (2l), (22), (23) (2001); NEB. REV. STAT. ANN. § 28-710 (Michie Supp. 2002); NEV. REV. STAT. ANN. §§ 432B.020, .065, .070, .090, .100, .110, .130, .140, .150 (Michie 2002); N.H. REV. STAT. ANN. § 169-C:3, (2002); N.J. STAT. ANN. § 9:6-8.9 (West 2002); N.M. STAT. ANN. § 32A-4-2 (Michie 1999); N.Y. SOC. SERV. LAW § 384-b(8)(a), (8)(b) (Consol. Supp. 2003); § 412(1), (2), (4)–(12); N.Y. FAMILY COURT LAW § 1012(e)–(h), (j) (Consol. Supp. 2003); N.C. GEN. STAT. § 7B-10l(1) to (3), (8), (9), (14), (15), (18), (19) (2001); N.D. CENT. CODE § 27-20-02(1) to (3), (8) (1999 & Supp. 2001); § 50-25.1-02; OHIO REV. CODE ANN. §§ 2151.011(B)(1), (22), (27)–(29), (33), (C), .03(A), (B), .031, .04, .05 (Anderson 1998 & Supps. 2000, 2002); § 2907.01(A)–(C); OKLA. STAT. ANN. tit. 10, §§ 7102(B), 7103(E), 7106(A)(3) (West Supp. 2003); OR. REV. STAT. § 419B.005 (1995); 23 PA. CONS. STAT. ANN. § 6303(a)–(b) (West 2001); R.I. GEN. LAWS § 40-11-2 (2002); S.C. CODE ANN. § 20-7-490(1)–(3), (4)–(21) (Law. Co-op. 1999); S.D. CODIFIED LAWS § 26-8A-2 (Michie 1999); TENN. CODE ANN. §§ 37-1-102(b)(1), -102(b)(12), -102(b)(21), -401, -602(a) (2001 & Supp. 2002); TEX. FAM. CODE ANN. § 261.001, (West 2002); TEX. PENAL CODE ANN. § 43.01 (West 1997); UTAH CODE ANN. § 62A-4a-402 (2000); VT. STAT. ANN. tit. 33, § 4912 (2001); VA. CODE ANN. § 63.1-248.2 (Michie 2002); § 63.2-100; WASH. REV. CODE ANN. §§ 26.44.015(1)–(3), .020(2)–(6), .020(12)–(16), .020(19), .030(l)(c) (West Supp. 2003); W. VA. CODE ANN. § 49-1-3(a) to (c), (e), (g), (h), (i) to (n), (q) (Michie 2001); WIS. STAT. ANN. §§ 48.02(1), (2), (2c), (4), (5j), (14g), .981(1) (1997 & West Supps. 1998, 2002); WYO. STAT. ANN. § 14-3-202(a)(i) to (a)(xi) (Michie Supp. 2002).

injuries will be identified and reported.[100] By 1967, the reporting of suspected child abuse was mandatory in all fifty states and the District of Columbia.[101] Since the early 1970s, reporting laws have expanded the statutory definition of child abuse, which often includes physical, sexual, emotional, and mental abuse as well as neglect and threat of future harm.[102] Currently, the Federal Child Abuse Prevention and Treatment Act[103] sets the following minimum standards for state definitions of child abuse as "any recent act or failure to act on the part of a parent or caretaker, which results in death, serious physical or emotional harm, sexual abuse or exploitation, or an act or failure to act which presents an imminent risk of serious harm."[104]

Legislatures have also expanded the range of professionals who are required to report suspicions of abuse. Every state and the District of Columbia have statutes identifying mandatory reporters,[105] which typically include: doctors, nurses, hospital personnel, dentists, medical examiners, coroners, mental health professionals, social workers, school personnel, law enforcement officials, and child care providers.[106] Recent scandals involving the sexual abuse of children by Catholic priests has focused national attention on the issue of clergy reporting requirements and prompted new legislation in several states.[107]

In cases involving physical abuse, medical experts typically become involved the moment a child with physical injuries is clinically evaluated in the emergency room or doctor's office. In certain cases, a preliminary abuse diagnosis will

---

[100] *See* NAT'L CLEARINGHOUSE ON CHILD ABUSE & NEGLECT, U.S. DEP'T HEALTH & HUMAN SERVS., CHILD ABUSE AND NEGLECT STATE STATUTE SERIES, COMPENDIUM OF LAWS, REPORTING LAWS, DEFINITIONS OF CHILD ABUSE AND NEGLECT 1 (2002), http://www.calib.com/nccanch/pubs/stats02/define.pdf [hereinafter STATE STATUTE SERIES] ("[R]eporting statutes . . . determine the grounds for State intervention in the protection of a child's well-being.") (footnote omitted).

[101] *See* Steven J. Singley, *Failure to Report Suspected Child Abuse: Civil Liability of Mandated Reporters*, 19 J. JUV. L. 236, 238 (1998).

[102] *See id.*

[103] 42 U.S.C. §§ 5101–5107, 5116–5116i (2000).

[104] 42 U.S.C. § 5106g(2).

[105] *See* STATE STATUTE SERIES, *supra* note 100, at 1.

[106] *See id.*

[107] *See, e.g.*, Michael Paulson, *Scandal Fallout: U.S. Bishops Vow Cooperation with Authorities on Sex Abuse*, BOSTON GLOBE, May 19, 2002, at A18 (describing how most state mandatory child abuse reporting statutes include clergy and how Catholic bishops will debate a proposed requirement that they report all abuse allegations to secular authorities regardless of the particular state statute); Carrie Budoff, *Clergy Bound by Two Laws: Statutes Conflict on Reporting Abuse of Children*, HARTFORD COURANT, May 28, 2002, at A1, LEXIS, Nexis Library, HTCOUR File (describing how Connecticut lawmakers are seeking to "lift the shroud of confidentiality surrounding the confessional" by requiring priests to comply with mandatory child abuse reporting statutes).

require some level of medical training and expertise.[108] In other cases, injuries such as burns or bruises on a small infant should raise the suspicions of the lay observer. When, for example, an infant presents with numerous broken bones, hospital staff must choose between one of only two possible diagnoses: (1) the child was abused[109] or (2) the child sustained some type of accidental trauma.[110]

There is also a remote possibility that the infant suffers from a rare disease, such as Osteogenesis Imperfecta ("OI"), which is sometimes referred to as "Brittle Bone Disease." Even an accurate diagnosis of OI does not establish the cause of the injury, but is instead a finding that might make certain injuries more consistent with accident than abuse. [111] Medical professionals and the courts must also bear in mind that children who suffer from bone disease may be equally likely, or even more likely, to be victims of abuse.[112] In the vast majority of cases, clinical assessments and laboratory tests enable physicians to exclude rare bone diseases.

When physicians suspect child abuse, they frequently refer to injuries as "suspicious" for abuse. A preliminary clinical finding of suspicious injuries instigates further efforts to confirm or refute this diagnosis. For example, when suspicious injuries include fractures, standard medical protocol requires that the child undergo a skeletal survey, which is a series of x-rays of the entire body.[113]

Judges should be aware that there are at least three types of radiological evidence that can confirm initial suspicions of abuse. First, doctors can find

---

[108] *See* Lyon et al., *supra* note 16, at 94 ("Determining whether a young child's injuries are due to physical abuse is often extremely difficult. Frequently, the child is nonverbal, and there are no witnesses other than the caretakers that are suspected of abuse. [This explains why] [e]xpert medical opinion is often necessary to diagnose abuse.").

[109] *See id.* at 102 ("Skeletal injury is a common manifestation of child abuse.").

[110] *See infra* Part V (discussing the medical literature that describes how to distinguish fractures attributable to abuse from those attributable to accident).

[111] *See* Jan Bays, *Conditions Mistaken for Child Abuse*, *in* CHILD ABUSE: MEDICAL DIAGNOSIS AND MANAGEMENT 358, 378 (Robert M. Reece ed., 1st ed. 1994) (noting that "[s]everal rare metabolic conditions are associated with bones that are easily fractured").

[112] Obviously, these explanations are not mutually exclusive. A child may, for example, have some type of bone disease *and* have been abused or suffered accidental trauma. "The possibility of intentional injury cannot be disregarded in osteogenesis imperfecta, since even children with this disease can be abused." Gregory D. Launius, *Radiology of Child Abuse*, *in* CHILD MALTREATMENT: A CLINICAL GUIDE AND REFERENCE 27, 51 (James A. Monteleone & Armand E. Brodeur eds., 1994); *see also* Bays, *supra* note 111, at 380 (discussing a case study involving a child diagnosed with both OI and abuse); Sheila Gahagan & Mary Ellen Rimsza, *Child Abuse or Osteogenesis Imperfecta: How Can We Tell?*, 88 PEDIATRICS 987, 988 (1991) (noting that child abuse could coexist with OI); Lyon et al., *supra* note 16, at 95 ("If accidents (such as a fall) and disease are ruled out, then physicians are confident in stating that the child's injury is 'nonaccidental,' that is, due to abuse.").

[113] Skeletal surveys can reveal additional fractures and can also help physicians to date the fractures enabling them to compare the radiographs to the history provided by the caretaker.

additional types of injuries suggestive of abuse, for example, subdural hemorrhages indicative of shaking. Second, doctors may discover additional fractures in various stages of healing, which indicate multiple injuries over some period of time. Third, doctors may conclude that a type of fracture is, itself, inconsistent with accidental trauma. Two examples of this third type of evidence, which will confirm an initial abuse diagnosis for many doctors, are metaphyseal fractures and rib fractures. "Metaphyseal fractures constitute highly suggestive evidence of abuse in the infant or young child when they occur in the area of bone between the metaphysis [the area where the bone flares out] and the epiphysis [the end of the bone]."[114] These "bucket-handle" fractures indicate that the limb was twisted or pulled.[115] Rib fractures are also "highly suggestive evidence of abuse in children under three years of age, particularly when located at the sides ('lateral') and back ('posterior') of the rib near the vertebral column, and, more rarely, when they involve the rib ends in the front ('anterior costochondral')."[116] Rib fractures in abused children are usually attributable to squeezing the chest,[117] but also can be caused by direct blows.[118]

Generally, while radiologic tests are being performed, clinical staff will compare the child's injuries with the history provided by the parent/caregiver.[119] Although child abuse may be diagnosed solely on the basis of unequivocal radiologic evidence of certain fractures or fractures that are not accompanied by a history of injury, an important factor in an abuse diagnosis may be any discrepancies between the history provided by the caretaker and the nature and extent of the child's injuries. "The correlation or lack of correlation of the diagnostic images with the temporal elements of the history is often the first evidence of suspicious trauma."[120]

---

[114] Lyon et al., *supra* note 16, at 109.

> Abusive metaphyseal fractures are often described as "bucket-handle" fractures or "corner" fractures. Such fractures are generally observed in infants or children under two years of age, whose metaphysis is more fragile than that of older children. Metaphyseal fractures can be caused by shaking as well as by pulling and twisting of the child's extremities.

Id.

[115] *See id.*; *see also* Gregory D. Launius et al., *supra* note 112, at 50 (noting that "[m]etaphyseal fractures are typically absent" in Osteogenesis Imperfecta); David F. Merten et al., *Skeletal Manifestations of Child Abuse*, *in* CHILD ABUSE: MEDICAL DIAGNOSIS AND MANAGEMENT, *supra* note 111, at 23, 48 (stating that "metaphyseal fractures rarely occur in OI").

[116] *See id.* at 209.

[117] *See id.* at 211 (describing rib fractures as "characteristic of abuse").

[118] *See* Lyon et al., *supra* note 16, at 111.

[119] Data indicate that 87.3% of all child abuse and/or neglect perpetrators are a parent of the victim. *See* CHILD MALTREATMENT 1999, *supra* note 26, at tbl.3-1.

[120] *Id.*

Even at this early phase of the investigative process, trained medical staff know that, whenever abuse is suspected, their findings may be subject to careful scrutiny by other physicians and, if the case proceeds to trial, by lawyers, judges, and jurors.[121]

> In a child abuse case it is particularly important to note all fractures and other osseous and soft tissue abnormalities and to date their occurrence if that is possible. Attorneys are especially interested in etiology, because a diagnosis of "accident" or "inflicted injury" may determine case disposition. Multiple fractures of different ages may permit a more convincing conclusion than a single, acute fracture, but the nature or location of the fracture (e.g., "corner" fracture, "bucket-handle" lesion, rib fractures) and the consistency between the injury and parental explanation/history are important data.[122]

This type of medical information is essential to both sides as they prepare for trial. It will shape the state's case and limit the possible defenses. This evidence will also be critical to the judge who must make admissibility determinations and the jurors who will evaluate the evidence at trial.

2. *Stage Two: Investigating the Physical Abuse of Children*

The second stage of a child abuse case involves the investigation of suspected abuse. During the investigative phase, the rights of the parent/caregiver are balanced against the obligation of the state to protect the child.

> The legal system is influential in establishing procedures to be followed in the investigation of a case, exceptions under which those procedures can be circumvented, and liability for officials who act outside the prescribed procedures. The standards of proof that are required to determine whether a case is founded (substantiated) or unfounded (unsubstantiated) are established by state legislatures and are presumed to have some impact on the caseworkers' or investigators' decision-making process.[123]

If child abuse has been detected and confirmed through investigation, the child abuse case will likely end up in court. Civil and criminal investigators often

---

[121] Medical professionals involved in the diagnosis of abuse can also serve as witnesses at trial. For example, physicians who testify based on their experience treating the patient or consulting with other physicians are sometimes referred to as "fact" witnesses, although in their testimony "they will also be applying medical expertise to a greater or lesser degree in assessing the significance of the patient's signs and symptoms." REFERENCE MANUAL, *supra* note 4, at 450.

[122] Richard Bourne, *Child Abuse and the Law*, *in* DIAGNOSTIC IMAGING OF CHILD ABUSE 243, 250 (Paul K. Kleinman ed., 1st ed. 1987).

[123] Repucci & Fried, *supra* note 26, at 107.

work closely with hospital staff to assess the viability of any legal proceeding. State and local agencies responsible for investigating cases of suspected child abuse will determine whether to proceed with a civil action (i.e., where the state's objective is to temporarily or permanently terminate parental rights). In these cases, "[c]ourts play a key role in determining whether children will be removed from their homes, how long they will remain in foster care, and where they will permanently reside."[124] These same agencies generally work in conjunction with state prosecutor's offices to decide whether a criminal case is warranted. More serious cases of abuse where the state has identified the defendant(s) are more likely to result in criminal charges.

3. *Stage Three: Adjudicating the Child Abuse Case*

Science, in the form of medical expert opinion, guides the adjudication of many child abuse cases. According to one trial judge, "in many such cases the only proof available is circumstantial evidence since abusive actions usually occur within the privacy of the home, the child is either intimidated or too young to testify, and the parents tend to protect each other."[125] Prior to trial, judges are frequently confronted with conflicting medical expert testimony. The government invariably offers medical evidence to establish that the child's injuries were caused by abuse. The defense often responds with medical expert testimony that supports a non-abusive explanation for the child's injuries.[126] "[T]he burden for the defense is often to prove the medical findings are not the result of shaking [or other physical abuse], but instead the product of some other disease process or circumstance, e.g. the child's injuries are consistent with an accidental fall."[127] The judge's decision to admit, exclude, or limit medical evidence offered by either side can have a significant impact on the outcome of the trial. In these cases, as in all cases with a powerful scientific determinant, judges must be especially careful not to admit scientific evidence that falls short of appropriate evidentiary standards.

Lawyers who prosecute or defend child abuse cases and judges who decide these cases inevitably become familiar with the diagnostic and investigative processes. A basic understanding of how child abuse is diagnosed can improve

---

[124] Victor E. Flango, *Measuring Progress in Improving Court Processing of Child Abuse and Neglect Cases*, 39 FAM. CT. REV. 158, 161 (2001).

[125] *In re* Sarah C., 626 N.W.2d 637, 643 (Neb. Ct. App. 2001) (quotations omitted).

[126] "It is difficult to imagine a child abuse case, whether it involves physical or sexual abuse, where the defense would not be aided by the assistance of an expert." Beth A. Townsend, *Defending the "Indefensible": A Primer to Defending Allegations of Child Abuse*, 45 A.F. L. REV. 261, 270 (1998).

[127] Brian K. Holmgren, *The Legal System's Role in Facilitating Irresponsible Expert Testimony*, SBS QUARTERLY, Summer 1999, at 1, *available at* http://www.dontshake.com/sbssummer99expert.html.

the quality of lawyers' legal argument and judge's decisions. However, the medical science of child abuse often remains "a mystery to many attorneys, even to those who routinely handle such cases."[128] This may be attributable to the fact that:

> [t]he medical literature is often impenetrable to those without special training, leading attorneys to defer to expert opinion without fully understanding the basis for such opinion. This is unfortunate. Without understanding the research that underlies expert medical judgment, an attorney can neither make full use of the physician's expertise, nor adequately cross-examine an opposing expert.[129]

This means that expert witnesses play a critical role in science-based child abuse cases. They review and explain complex evidence. Experts draw inferences from the existing body of medical literature and apply causation theories to the facts at issue. They assess how well the evidence supports both sides of the case and identify contradictions and weaknesses. Expert witnesses also help the attorneys prepare for direct and cross examination, and in perhaps their most important role, experts can testify to their theories, opinions, and conclusions at trial.[130]

Although judges and jurors may need expert assistance to decide certain factual and legal questions, experts can sometimes do more harm than good. In a recent article in *Pediatric Radiology*, the official journal of the Society for Pediatric Radiology, two pediatric radiologists, Doctors Stephen Chapman and Christine Hall, stated that experts who testify in child abuse cases must remain cognizant of certain professional and ethical obligations.

> Experts must not put forward untested or unacceptable views, and must be prepared to cast aside ideas of loyalty to one party or another and give evidence with the child's welfare as the primary aim. Conclusions must be reached after considering all the available evidence, not just those aspects which support a particular view.[131]

Doctor Paul Kleinman, a pediatric radiologist and world-renowned expert on child abuse, has described a different dynamic that may prevent medical experts from adequately assisting judges and juries.

> [P]hysicians and other professionals are often reluctant to make clear-cut and unequivocal statements in a courtroom and are much more likely to give "truer"

---

[128] Lyon et al., *supra* note 16, at 94.

[129] *Id.*

[130] *See* FED. R. EVID. 702 (permitting expert witnesses to testify in the form of an opinion).

[131] Stephen Chapman & Christine M. Hall, *Non-Accidental Injury or Brittle Bones*, 27 PEDIATRIC RADIOLOGY 106, 109 (1997).

> readings informally and off the record. Although clearly one should never present opinions unjustified by the data, it is essential to avoid overcaution. When a child appears with characteristic radiologic patterns of abusive injury, and abuse is the only reasonable explanation, the expert should so state that under oath in the courtroom.[132]

Although experts who cannot state a clear medical opinion provide little meaningful assistance to the trier of fact, judges should be skeptical of experts who offer definitive conclusions that are inadequately supported by valid scientific evidence.

Judges confronted with expert opinions derived from complex medical data should avoid the temptation to defer to the scientist because this has the effect of relaxing the admissibility standard. In particular, courts may be willing to grant too much latitude to medical experts in civil child protection proceedings.[133]

> Because the best interests of the child rather than the criminal responsibility of the adult are at issue, courts may find testimony about the results of clinical evaluations or clinical opinions particularly helpful. In addition, because the standard of proof in civil proceedings is lower, the courts may consider this standard to be more comparable with the criteria used in clinical decision making. Of course, opposing lawyers are free to challenge the basis for the opinion, argue for alternative explanations, or point out the limitations of clinical determinations.[134]

A related concern is that in criminal cases:

> [j]udges are more likely to admit a defense expert than to exclude them [sic]. While judges exercise broad discretion in determining the admissibility of expert testimony, judges are much more likely to exercise that discretion to exclude a government expert than a defense expert. This results in part from a judicial philosophy that favors the rights of defendants to present a defense, and a concomitant fear by trial judges that if they exclude a proffered defense expert they may be reversed on appeal.[135]

The only way for judges to avoid infecting the legal analysis with unreliable medical expert evidence is for courts to strictly adhere to the gatekeeping role. This may require that judges amass information about existing scientific controversies, which may effect their reliability assessments.

---

[132] *Bourne*, *supra* note 122, at 250.

[133] *See* Lucy Berliner, *The Use of Expert Testimony in Child Sexual Abuse Cases*, *in* EXPERT WITNESSES IN CHILD ABUSE CASES: WHAT CAN AND SHOULD BE SAID IN COURT 11, 15 (Stephen J. Ceci & Helene Hembrooke eds., 1998).

[134] *Id.*

[135] Holmgren, *supra* note 127, at 1.

V. CONTROVERSIAL SCIENTIFIC EVIDENCE: WERE THE CHILD'S INJURIES CAUSED BY ABUSE OR ACCIDENT?

> [I]t is only by examining scientific controversies while they are in progress that the mechanism by which ships (scientific findings) get into bottles (validity) can be understood. If this process is not seen in operation it may be thought that ships were always in bottles, and that all scientists did was find them ready assembled, as it were.
>
> Harry M. Collins[136]

A. *Osteogenesis Imperfecta and Accidental Injuries*

A complex medical/legal question has arisen in cases where infants present with multiple fractures and child abuse is suspected. Judges frequently admit defense medical evidence concluding that these infants have not been abused, but instead suffer from a variant form of Osteogenesis Imperfecta ("OI"). This theory is offered in cases where the medical evidence of abuse involves fractures in different stages of healing, the infants have tested negative for conventionally diagnosable metabolic bone diseases, and the fractures do not continue to occur after the child has been placed in protective custody.[137] At trial, the defense relies on medical expert testimony to argue that the child was not abused, but instead suffers from a mild and previously undiagnosed form of OI.

The first step in the admissibility inquiry should be for the court to place the diagnosis in context. To assess the validity of this medical evidence, judges should distinguish testimony about OI, a rare but well-recognized disease, from testimony about variant forms of OI, which are novel and controversial.

Osteogenesis Imperfecta has been described in the medical literature as an "inherited disorder of connective tissue with deficiency of type I collagen leading to abnormal bone formation and increased bone fragility. As a result, trivial injuries may cause fractures in these patients."[138] The legal significance of OI in child abuse cases has been specifically recognized by medical experts. "Of all the various conditions invoked by parents and their legal representatives to explain inflicted fractures, OI is cited most frequently. It is therefore essential to be

---

[136] Edmond & Mercer, *supra* note 14, at 39 (quoting Harry M. Collins, *The Seven Sexes: A Study in the Sociology of a Phenomenon or the Replication of Experiments in Physics*, 9 SOC. 205 (1975)).

[137] *See* Daniel R. Cooperman & David F. Merten, *Skeletal Manifestations of Child Abuse*, *in* CHILD ABUSE: MEDICAL DIAGNOSIS AND MANAGEMENT, *supra* note 111, at 123, 146 (noting that OI is the medical condition most frequently advanced as a defense to child abuse allegations).

[138] *Id.* at 149.

familiar with the classification of OI and the features that distinguish it from child abuse."[139]

Osteogenesis Imperfecta has been classified into four major types, depending on the age of onset of fractures, extraskeletal manifestations, and mode of inheritance.[140] Infants with types I and II OI account for 80% of all cases of OI.[141] The medical literature indicates that infants suffering from types I and II OI should never be confused with patients suffering from abuse.[142] This is because OI has obvious clinical manifestations. For example, all children with type I or II OI have blue sclerae (i.e., the white area of the eye looks blue). In addition, type II OI is almost invariably lethal in the perinatal or neonatal period.[143] Of the remaining 20% of children who suffer from OI types III or IV, those who have type III should have wormian bones and osteoporosis, which are easily detected through radiologic tests.[144] These clinical findings, along with other readily identifiable features of OI types I, II, and III, enhance the likelihood that physicians examining a patient for OI can make a reliable diagnosis.

The only type of OI that might be confused with child abuse is type IV, which tends to be the least severe form of OI. However, even type IV has certain clinical indicators that experts should consider during the diagnostic process.

> Patients with OI type IV have variable degrees of short stature, with mild to moderate deformity. Fractures may begin to occur prenatally and may be associated with deformity of the long bones that is evident at birth. . . . Affected patients generally have a triangular head, with a prominent forehead. Sclera are generally normal, except in infancy, when they have a blue hue. . . . Radiologic examination demonstrates osteoporosis, mild to severe bowing of the long bones, and spinal deformity.[145]

Doctor Deborah Ablin, a pediatric radiologist with expertise in child abuse has opined that the likelihood of a clinician seeing a child with mild type IV OI and with white sclerae, normal hearing, normal dentition, negative family history, and no wormian bones is exceedingly rare.[146] According to Dr. Ablin, if this "unusual case" does occur, a definitive diagnosis can be made using collagen

---

[139] *Id.*

[140] *See* Lachman et al., *supra* note 20, at 197–98.

[141] *See* Cooperman & Merten, *supra* note 137, at 149.

[142] *See id.*

[143] *See id.*

[144] *See id.*

[145] Arthur Zinn, *Genetic Disorders That Mimic Abuse or SIDS*, *in* CHILD ABUSE: MEDICAL DIAGNOSIS AND MANAGEMENT, *supra* note 111, at 404, 412.

[146] Deborah S. Ablin, *Osteogenesis Imperfecta: A Review*, 49 CAN. ASS'N RADIOLOGISTS J. 110, 111 (1998).

analysis.[147] Thus, when a judge is assessing the reliability of a diagnosis of OI of types I–IV, the medical evidence indicates that it is "generally uncomplicated to distinguish OI from child abuse."[148]

B. *Evaluating Variant Forms of a Well-Recognized Disease: Temporary Brittle Bone Disease*

The data on OI types I–IV must be contrasted with the data describing variant forms of OI, such as Temporary Brittle Bone Disease ("TBBD").[149] When evidence of variant forms of a disease is proffered, validity assessments necessarily involve some understanding that new diseases are difficult to diagnose. This is attributable, in part, to the lack of an accepted methodology for recognizing new diseases and to

> a relatively small number of discrete symptoms and signs [that] are shared by a much larger number of coherent diseases. . . . [W]hen a possible new set of characteristic symptoms, signs, and laboratory manifestations is described, there is no one method for developing consensus on whether a new disease entity exists. For example, when the characteristic symptoms, signs, and laboratory test results of acquired immunodeficiency syndrome (AIDS) were first described in the early 1980s, prior to the identification of the human immunodeficiency virus (HIV), there was considerable controversy over whether a new disease entity had manifested itself. Development of the test for infection with the specific virus cemented recognition of the disease.[150]

TBBD was first described in 1990 at the Fourth Annual Conference of Osteogenesis Imperfecta,[151] as a short-lived developmental bone disease that results in easy bone fracturability in very young children for a limited period of time.[152]

---

[147] *Id.*

[148] *See id.*; *see also* Ablin & Sane, *supra* note 24, at 111 ("[I]n the majority of instances, the correct diagnosis can be reached by careful appraisal of social and family history and careful clinical and roentgenographic examination.").

[149] *See, e.g.*, Ralph Hicks, *Relating to Methodological Shortcomings and the Concept of Temporary Brittle Bone Disease*, 68 CALCIFIED TISSUE INT'L 316, 316–19 (2001); Mark E. Miller & T.N. Hangartner, *Temporary Brittle Bone Disease: Association with Decreased Fetal Movement and Osteopenia*, 64 CALCIFIED TISSUE INT'L 137, 137–42 (1999); Colin R. Paterson & Susan J. McAllion, *Osteogenesis Imperfecta in the Differential Diagnosis of Child Abuse*, 299 BRIT. MED. J. 1451, 1451–54 (1989).

[150] REFERENCE MANUAL, *supra* note 4, at 462.

[151] *See* Ablin & Sane, *supra* note 24, at 111.

[152] *See* Ablin, *supra* note 146, at 110–23 (noting that TBBD "remains a medical hypothesis lacking the support of sound scientific data").

Recently the Arizona Supreme Court addressed the question of whether it was an abuse of the trial court's discretion to exclude defense testimony on TBBD.[153] In this case, doctors discovered and then reported numerous suspicious fractures in a three-month-old girl.[154] Following a state investigation, the girl's mother was charged with eleven counts of child abuse.[155] The defense relied on the medical theory that the fractures were not abuse, but were the result of TBBD. After numerous hearings on whether the defense could introduce testimony on TBBD,[156] the trial judge excluded testimony by the world's leading TBBD expert, Dr. Colin Paterson. The trial court ruling was a sanction for late disclosure to the prosecutor and did not address the question of reliability.[157]

The Arizona Supreme Court found that the trial court had clearly abused its discretion. Interestingly, the supreme court went well beyond the scope of the pretrial ruling and made very specific findings about the *reliability* of the excluded scientific evidence. The court began with the assumption that TBBD is a reliable medical diagnosis, referring to Dr. Paterson's status as "arguably the world's preeminent TBBD expert."[158] The court then concluded that Dr. Paterson's testimony should have been admitted under the state's "general acceptance" standard, noting that "[t]he value of his testimony was . . . clear to everyone."[159] The court also assumed that Dr. Paterson, if he had been permitted to testify, would have successfully impeached the prosecutor's expert, "something he was capable of doing based in part on his vast experience in defining and diagnosing TBBD."[160] Finally, the court concluded that the testimony of the prosecutor's witness criticizing TBBD certainly would have been undermined because "[h]ad [Dr.] Paterson testified, he would have discussed the array of cases he has seen in light of his own experience and the diagnoses arising therefrom."[161] Based on these assumptions, the court held that the exclusion of Dr. Paterson's TBBD evidence "deprived the defendant of the only real opportunity she might have had to introduce meaningful exculpatory evidence."[162]

The Arizona Supreme Court creates an imprimatur of scientific reliability that should be carefully evaluated. First, the court has made a global assessment of the

---

153 State v. Talmadge, 999 P.2d 192 (Ariz. 2000).

154 *Id.* at 193.

155 *Id.*

156 *Id.* at 193–95.

157 *Id.* at 194–95.

158 *Id.* at 194.

159 *See Talmadge*, 999 P.2d at 196 n.5.

160 *Id.* at 196.

161 *Id.*

162 *Id.* at 197.

validity of TBBD as a disease.[163] Second, by highlighting the relevance and reliability of Dr. Paterson's diagnosis in *this* case, the court speculates that this scientific evidence was reliably applied to the facts at issue. The court reaches these conclusions based in large measure on Dr. Paterson's credentials.

This case should be contrasted with the resistance that TBBD has encountered in England. After a lengthy hearing addressing the admissibility of TBBD evidence, Judge Peter Singer of the Royal Courts of Justice, Family Division found this evidence not only inadmissible, but scientifically invalid.

> I have dealt with these matters at such length in an attempt to demonstrate what in my judgment is the subjectivity, the unreliability, the unscientific and unproved nature of Dr. Paterson's speculations that TBBD exists as a clinical entity, and (in particular) that X in any event falls within the syndrome. It is, in my opinion, a syndrome which can only be recognized by someone with tunnel vision who notes only those positive factors which are self-selected, and adapts his description of the disease as he goes along, thus enabling him to disregard, indeed to ignore, factors which from his own published work one would suppose he might regard as relevant.[164]

Judge Singer concluded with a highly critical assessment of TBBD and Dr. Paterson's methodology. "In my judgment, in relation to any future potential diagnosis by Dr. Paterson of TBBD, his methodology and his credentials to express opinion deserve to be and should be subjected to rigorous scrutiny before he is given leave to report in further cases."[165]

VI. ASSESSING THE VALIDITY OF NOVEL SCIENTIFIC EVIDENCE: THE USE AND APPLICATION OF A MORE FOCUSED INQUIRY

If a large number of judges is clearly confused or ill-informed about basic scientific principles, *Daubert* cannot be accurately or consistently applied. The result is that a hodgepodge of reliable and unreliable scientific evidence is entering our courtrooms. There are short and long term solutions to this problem. In the long term, judges could work together with scientists and other experts to enhance their scientific sophistication and thereby improve the accuracy of legal

---

[163] Although the *Kumho* decision clearly indicates that global assessments of reliability are not required, *see* Moreno, *supra* note 27, at 1055; Risinger, *supra* note 89, at 773, Arizona's continued reliance on a general acceptance standard could result in judicial assessments of general reliability.

[164] *Re* X (Non-Accidental Injury: Expert Evidence), 2 F.L.R. 1, 27 (Royal Courts of Justice, Fam. Div. 2001).

[165] *Id.*

decisions involving scientific evidence.[166] In the short term, judges could adopt a new approach to the task of assessing scientific evidence.

> Judges need to know what critical questions to ask, they need to know what methodological and statistical issues scientific experts, and other purveyors of science, should address and comment on when proffering science for use in the court. Judges need to know what to listen and look for when expert evidence is presented and what they should be asking about when the information is not forthcoming.[167]

According to the Supreme Court, the law grants a judge "broad latitude when [she] decides *how* to determine reliability."[168] However this goal may have been undermined by the *Daubert* court's inclusion of four (reliability) markers, which have for the past decade been mechanically applied without yielding much insight. It is time for courts to return to the original concept of flexibility and think creatively about how to improve the process of assessing scientific evidence. We might begin this process with the following question:

> In regard to any proffer of expertise, is there good reason to believe that the proffered product of the claimed expertise (given its specific form and methods and condition of which it is a product) provides the jury with appropriately reliable information on the case specific question upon which the expert is proffered?[169]

A first step toward answering this question might be to develop a simple structure involving basic questions for judges to ask experts.[170] Questions could be crafted so that the judge moves chronologically through the scientific process exposing, in order and where appropriate, the scientist's hypothesis, data collection techniques, methodology, conclusion, and elimination of alternative conclusions. An inquiry of this type would provide structure, help clarify the elements of the dynamic process of developing scientific conclusions that bear on

---

166 *See* Moreno, *supra* note 27, at 1087–91.

167 Gatowski et al., *supra* note 7, at 455.

168 Kumho Tire Co. v. Carmichael, 526 U.S. 137, 142 (1999) (citing Gen. Elec. Co. v. Joiner, 522 U.S. 136, 143 (1997)). The Supreme Court has also been explicit that neither Rule 702 nor *Daubert* should be rigidly interpreted or applied. According to the *Kumho Tire* Court, the pretrial inquiry under Rule 702 is "flexible," and the *Daubert* factors "do *not* constitute a 'definitive checklist or test.' " *See Kumho Tire*, 526 U.S. at 150 (citing Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 594 (1993)).

169 *See* Denbeaux & Risinger, *supra* note 28, at 19.

170 Judges have broad discretion to fashion the validity inquiry. *See, e.g.*, *Kumho Tire*, 526 U.S. at 152; Goebel v. Denver & Rio Grande W. R.R., 215 F.3d 1083, 1087 (10th Cir. 2000). *Kumho Tire* recommends that judges adopt a practical and flexible inquiry that considers the relevant factors in the circumstances of the case. *See Kumho Tire*, 526 U.S. at 149–52.

validity, expose methodological flaws, avoid unwanted inferences, and clarify the legal standard.[171]

The following three questions illustrate a very simple example of this approach, which is not only appropriate in a child abuse case context but also could be modified to fit the type of evidence under consideration. The judge could use the pretrial hearing to answer these questions:

(1) How did the experts arrive at their conclusions?

(2) How did the experts test their conclusions?

(3) How did the experts rule out other conclusions?

Questions of this type could help judges determine whether the testimony satisfies the Rule 702 requirements that "[(1)] the testimony is the product of reliable principles and methods, and . . . [(2) that] the witness has applied the principles and methods reliably to the facts of the case."[172] Although I have written elsewhere that *Kumho* and the 2000 amendments to Rule 702 reflect a move from global to local reliability,[173] I agree that local reliability alone is insufficient to warrant admission. "[E]xpert testimony cannot advance accurate outcomes locally unless it rests on acceptable epistemological warrant globally. A necessary but not sufficient condition of appropriate testimony 'locally' is reliable expertise 'globally.' "[174] Thus, judges must explore the validity of an expert's theory or technique and also evaluate the specific inferences and conclusions that the expert intends to advance in this case.[175]

What follows is a brief demonstration of the effect of posing these three questions to medical experts who proffer novel disease variants in child abuse cases.

A. *First: How Did the Experts Arrive at Their Conclusions?*

If judges begin by focusing on how experts arrive at their conclusions, they can elicit a substantial amount of information essential to a determination of global and local scientific validity. In general, judges should start with a broad perspective on the scientific question at issue, so that the experts' testimony can be understood in context. A second step is for the courts to narrow their focus to the accuracy of the experts' methodology in this case.

---

171 "Where the proffered expert offers nothing more than a 'bottom line' conclusion, he does not assist the trier of fact." Clark v. Takata Corp., 192 F.3d 750, 759 (7th Cir. 1999) (citing Rosen v. Ciba-Geigy Corp., 78 F.3d 316, 318–19 (7th Cir. 1996)).

172 FED. R. EVID. 702.

173 *See* Moreno, *supra* note 27, at 1052–55.

174 Allen, *supra* note 22, at 3. "Simply put, no matter how well credentialed and conversant in an established field, an expert may still testify to falsehoods. These falsehoods may involve generalities of the substantive content of the relevant field or its methodology, or as either applies to the particular facts of the case at hand." *Id.*

175 *See Daubert*, 509 U.S. at 594.

1. *Establishing Context: How Disease Prevalence Affects the Accuracy of Medical Diagnosis*

Every scientific conclusion must be understood in context. For example, it is impossible to assess the accuracy of any medical diagnosis unless the court begins with an understanding of disease prevalence. This is a basic statistical assumption. The more common any disease is in a given population, the more likely it is that a diagnosis of the disease is accurate.[176] This concept is sometimes expressed in the colorful truism that if you hear hooves behind you, you are more likely to encounter a horse than a zebra.[177] This means that in child abuse cases judges should begin by asking questions that will enable them to compare the relevant population base rates for each competing diagnosis. It may be particularly important for judges to incorporate base rate information into their pretrial screening process because "people are not good at integrating base rate information into their reasoning."[178]

a. *Applying This Inquiry to the Child Abuse Model: What Is the Base Rate for Child Abuse?*

Child abuse is a relatively common phenomena. Recent data indicate that in 2000 there were 879,000 child victims of physical abuse and neglect

---

[176] *See* FOSTER & HUBER, *supra* note 1, at 114.

[177] Thomas Bayes was the first person to create a mathematical theorem designed to incorporate information about base rates and observational accuracy. Bayes' theorem and calculations, first published in 1763, address the probability of causes, framing the question as follows: "Given that an event that may have been the result of any of two or more causes has occurred, what is the probability that the event was the result of a particular cause?" In general, although we cannot infer specific causation based on probabilities, this information is still highly relevant to our causation determinations. *See generally* Richard Overstall, *Mystical Infallibility: Using Probability Theorems to Sift DNA Evidence*, 5 APPEAL 28 (1999) (describing Bayes Rule).

[178] Lawrence M. Solan & Peter M. Tiersma, Language on Trial 12 (unpublished manuscript, on file with the author). Professors Solan and Tiersma note that the classic demonstration of this problem was made by psychologists Amos Tversky and Daniel Kahneman in 1982. *See id.* (citing Amos Tversky & Daniel Kahneman, *Evidential Impact of Base Rates*, *in* EVIDENTIAL IMPACT OF BASE RATES: HEURISTICS AND BIASES 156–57 (Daniel Kahneman et al. eds., 1982)). In this research, subjects were asked to evaluate the probability that a cab involved in an accident was blue, rather than green. Subjects were told that 85% of the cabs in the city are green and 15% are blue and that witnesses correctly identify the cabs 80% of the time. Given this information, most people will respond that witnesses will accurately identify a blue cab as blue 80% of the time. This is a vast overestimation of accuracy that fails to take base rates into account. In fact, a witness who identifies a blue cab as blue will only be accurate twelve out of twenty-nine times, or 41% of the time. Solan & Tiersma, *supra*, at 12–13 (citing Tversky & Kahneman, *supra*, at 156–57).

nationwide.[179] The number of reported child abuse fatalities for the same period was 1,200.[180] The highest victimization rate is in the 0 to 3 age group.[181] Although it is difficult to gauge the reliability of any statistical estimate, there is widespread agreement that child abuse, particularly abuse fatalities, is vastly under reported.

> Estimates regarding the incidence of child abuse are fraught with difficulties. They center around variations in reporting procedures, as well as different attitudes regarding what constitutes child abuse. It is safe to assume that official statistics simply reflect the workings of a system as it attempts to characterize the magnitude of the problem. The numbers do not provide a true measure of the extent of abuse and violent behavior in families.[182]

An additional problem inherent to the compilation of abuse statistics is that there is no general agreement on a definition of child abuse. "As the concept of child abuse has broadened to include sexual assaults, neglect, and adverse psychological consequences of abnormal family interactions, it is apparent that definitions will vary depending upon the professionals involved and the particular population and problem under study."[183] This means that normal statistical variations may be exacerbated by differences attributable to inconsistent definitions of abuse.

b. *What Is the Base Rate for the Proffered Diagnosis?*

Judges assessing the validity of any medical diagnosis should begin with questions designed to reveal the disease base rate. For example, when a defense expert has diagnosed the recognized disease of Osteogenesis Imperfecta in a case where a child's injuries are also consistent with abuse, judges should first understand that OI is extremely rare.

> Osteogenesis imperfecta (OI) is a rare genetic disorder characterized by bone fragility and frequent fractures. Infants and occasionally toddlers presenting with unexplained fractures may have OI raised as a possible explanation for the

---

[179] *See* SUMMARY 2000, *supra* note 16.

[180] *Id.*

[181] *See id.* (citing statistics indicating that most abuse victims are younger than three years old).

[182] Paul K. Kleinman, *Introduction*, *in* DIAGNOSTIC IMAGING OF CHILD ABUSE, *supra* note 122, at 2.

[183] *Id.* at 1.

> fractures. OI is a rare condition with an incidence estimated to be between 1 in 15,000 to 1 in 60,000 births. Child abuse is much more common.[184]

Of course, rarity alone does not mean that OI should not be considered by the physician as part of the diagnostic process or that this diagnosis is inaccurate. "Physicians seeing a child with unexplained fractures should consider OI when appropriate. Nonetheless, the frequency of child abuse is orders of magnitude greater than OI, especially for subtle cases of OI."[185]

In addition to understanding the base rate for OI, judges and lawyers must understand and distinguish base rates for variant forms of OI, such as Temporary Brittle Bone Disease ("TBBD"), where the normal symptomatology for OI is absent. In child abuse cases, defense experts

> often assert[ ] that a child suffers from a Type IV form of OI, in which the child may be presented without blue sclerae, wormian bones, osteoporosis, or a family history of OI. It has been estimated that the chance that fractures in a child under one are attributable to such a case is only one in three million.[186]

An appropriate first question for a court confronted with a diagnosis of a novel variant of OI might be how often a child will suffer from a variant of OI and have no signs or symptoms of the disease beyond multiple fractures. A recent study "calculated the probability of encountering a child under 1 year [of age] with OI and no other features or family findings of the disease as between one in 1 million and one in 3 million, or an annual incidence of one case every 100 to 300 years in a city of half a million people."[187] In the words of one medical expert, "[g]iven the rarity of this type of OI (1:1 to 3 million births) . . . relative to the frequency of child abuse, the probability of [diagnostic] error is minimal."[188] These data provide some sense of the possible base rates for variant forms of OI.

---

[184] Leonard E. Swischuk, *Radiographic Signs of Skeletal Trauma*, *in* CHILD ABUSE: A MEDICAL REFERENCE 151, 170 (Stephen Ludwig & Allan E. Kornberg eds., 2d ed. 1992).

[185] *Id.* at 171.

[186] Lyon et al., *supra* note 16, at 126.

[187] Jan Bays, *Conditions Mistaken for Child Physical Abuse*, *in* CHILD ABUSE: MEDICAL DIAGNOSIS AND MANAGEMENT 177, 200 (Robert M. Reece & Stephen Ludwig eds., 2nd ed. 2001).

[188] Cooperman & Merten, *supra* note 137, at 149.

2. *How Reliable Is Any Medical Diagnosis: What Is the Relationship Between General and Specific Causation?*

a. *Distinguishing Between General and Specific Causation*

Judges must sometimes assess the validity of the diagnosis of a single patient, also known as a determination of "specific causation." As we all know, doctors routinely make diagnoses for the purpose of providing treatment. This process of evaluating possible causal factors is sometimes referred to as differential diagnosis, a term "most physicians use . . . to describe the process of determining which of several *diseases* is causing a patient's *symptoms*."[189] Courts often use this term in a slightly different way to "describe the process by which causes of patient's condition are identified, particularly causes *external* to the patient."[190] One federal judge, describing the general view of most courts, defined this diagnostic process as a "methodology [that] is used time and again on a daily basis by medical doctors in diagnosing and determining the cause of illnesses in their patients, and . . . such methodology is a reliable basis for expert testimony regarding specific causation."[191] Generally, once a doctor has a working assumption regarding specific causation (for example, a patient's sore throat was caused by *streptococcus*) and treatment has begun, the causation inquiry is complete, unless something happens to indicate that the initial determination was inaccurate.[192]

The diagnostic process of distinguishing between child abuse and OI was described in the medical literature in a 1997 article in *Pediatric Radiology*. Two pediatric radiologists, Doctors Stephen Chapman and Christine Hall, described the factors doctors should consider in these cases when making a specific causation determination.[193] Although this article was designed to guide physicians who treat children and may need to distinguish between OI and child abuse, medical articles of this type, which detail diagnostic criteria, can be immeasurably helpful to judges who must assess the validity of medical expert testimony. These doctors suggested that whenever physicians are confronted with an infant with multiple fractures they must decide:

> (1) If there is truly an injury and not one of the many developmental variants which are so common in children.

---

189 REFERENCE MANUAL, *supra* note 4, at 443.

190 *Id.* at 443–44 (emphasis added).

191 Smith v. Pfizer, Inc., No. CIV.A.98-4156-CM, 2001 WL 968369, at *9 (D. Kan. Aug. 14, 2001).

192 *See* REFERENCE MANUAL, *supra* note 4, at 468 (describing how doctors routinely make decisions regarding diagnoses of illness and specific causation).

193 *See* Chapman & Hall, *supra* note 131, at 106.

> (2) Whether the explanation is appropriate for the injury sustained, i.e., an accident.
> (3) Whether the explanation is inappropriate in terms of mechanism, force or dating of the injury, i.e., there must be a suspicion of NAI [non-accidental injury] or fragile bones. [And]
> (4) If there is evidence of an underlying skeletal abnormality which has predisposed to the fracture, i.e., there are fragile bones.[194]

They also cautioned physicians that these diagnoses are not necessarily mutually exclusive because "children with medical conditions can be and have been abused and this includes children with osteogenesis imperfecta."[195] Ultimately, these experts suggest that a determination of specific causation should be based, in the "vast majority of cases[, on] the clinical findings, family history and radiological appearances [that] will differentiate between NAI, accidental trauma and fractures resulting from inappropriate trauma because of fragile bones."[196] It is easy to see how these criteria could form the basis of a judicial inquiry into how medical experts for either side arrived at their conclusions.

Judicial inquiries into the reliability of medical expert testimony must also involve some understanding of the relationship between general and specific causation. An opinion on specific causation identifies a particular factor that produced an identified result (e.g., this patient's sore throat was caused by *streptococcus*). Theories of general causation posit that these factors will produce similar results across a group or population (e.g., how many people, similar to the patient in age and other physical characteristics, in a given population will have sore throats caused by *streptococcus*).[197] "General causation is established by demonstrating, often through a review of scientific and medical literature, that exposure to a substance can cause a particular disease . . . ."[198] Theories of specific causation necessarily involve principles of general causation which must themselves be valid. Although medical experts can offer opinions on both general and specific causation because of the nature of legal questions, they are most commonly asked for specific causation conclusions.[199] As Professor David Faigman has described, this in itself can be problematic:

---

194 *Id.* at 106.

195 *Id.* "Unfortunately, children with OI may also be abused. In such cases, the diagnosis of child abuse may be made on the basis of fracture patterns typical of inflicted injury that are inconsistent with the history and findings of the physical examination." Cooperman & Merten, *supra* note 137, at 149–50.

196 Chapman & Hall, *supra* note 131, at 106.

197 *See* 1 DAVID L. FAIGMAN ET AL., MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY 32–34 (2002) (defining general and specific causation).

198 REFERENCE MANUAL, *supra* note 4, at 444.

199 *See id.* at 444–45 (noting that experts are more likely to testify about specific causation as it relates to an individual patient's medical condition).

> Presumably, medical doctors' experience gives them insights not shared by the average trier of fact or judge. Should the courts admit this experience in the form of expert testimony? The answer is, it depends. In science, experience usually is where the process begins, not ends. Experience provides insights useful for generating hypotheses that can be tested more systematically and more rigorously. It might be, for instance, that clinical experience indicates a relationship between silicone implants and autoimmune disorders. But the scientific arsenal contains a battery of weapons that can be brought to bear on this question, methods that have far greater power than the relatively myopic perspective of casual observation.[200]

Because *Kumho* extends the requirement that judges evaluate scientific evidence to non-empirical and experience-based methodologies, judges must assess the reliability of medical expert testimony based on clinical examinations or general clinical experience.[201] Medical diagnoses for treatment purposes are not systematically or rigorously tested. This means that courts should be cautious about accepting medical opinions as reliable scientific evidence when the data on general causation does not support the diagnosis (that is, the specific causation conclusion) or when the data on general causation is widely disputed.[202]

In *Moore v. Ashland Chemical, Inc.*,[203] the Fifth Circuit, applying *Daubert*, upheld the district court's exclusion of a medical expert opinion on the specific cause of plaintiff's disease. The court found that medical expert testimony cannot be admitted unless there can be some "objective, independent validation of the expert's methodology [and that t]he expert's assurances that he has utilized generally accepted scientific methodology [are] insufficient."[204] This is consistent with the argument that "with no proof of general causation, an expert should not be permitted to testify about specific causation,"[205] and the conclusion that "[w]ithout global reliability, one has gibberish."[206]

---

[200] DAVID L. FAIGMAN ET AL., MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY § 1-3.5 (1999).

[201] Kumho Tire Co. v. Carmichael, 526 U.S. 137, 148–49 (1999).

[202] *See* Ablin & Sane, *supra* note 24, at 111 (describing how none of the postulates used to support courtroom diagnoses of TBBD "have been substantiated . . . with sound scientific data or subsequently corroborated in any peer-reviewed journals . . . with prospective scientific research data"); Chapman & Hall, *supra* note 131, at 107 (describing how medical studies can undermine the assumption that OI and child abuse are commonly confused because studies indicate that children who were misdiagnosed as abuse victims when they had OI had either clinical/radiologic indicator of OI or a suggestive family history).

[203] 151 F.3d 269 (5th Cir. 1998) (en banc).

[204] *Id.* at 276 (citation omitted).

[205] FAIGMAN ET AL., *supra* note 67, at 32.

[206] Allen, *supra* note 22, at 6.

b. *General Causation: What Studies Has the Expert Relied Upon?*

Doctor Colin Paterson, widely acknowledged to be the most prominent expert on Temporary Brittle Bone Disease has testified in over one hundred child abuse cases and has published articles describing how to differentiate between TBBD and child abuse.[207] In an influential 1989 article entitled *Osteogenesis Imperfecta in the Differential Diagnosis of Child Abuse*,[208] Dr. Paterson and Dr. Susan McAllion studied 802 patients diagnosed with osteogenesis imperfecta. In 96 of these OI cases, nonaccidental injury had been suspected and in 15 child abuse investigations had been initiated.[209] Doctors Paterson and McAllion posited that none of the 802 cases involved abuse.[210]

Judges confronted with this study, or testimony based on this study, even in a *Daubert* jurisdiction, might not think to question the validity of these experts' conclusions. Even conscientious judges taking their gatekeeping responsibilities to heart might not know how to evaluate these published findings. These judges might begin by ascertaining how other physicians have assessed the scientific methodology used to generate these conclusions. In the pediatric textbook, *Child Abuse: Medical Diagnosis and Management*, Dr. Jan Bays evaluated the validity of the empirical research published by Dr. Paterson. Doctor Bays initially expressed concern about the diagnostic criteria that Dr. Paterson used in his two published studies.[211] Doctor Bays specifically noted that Dr. Paterson "postulated that these children had either OI or a new entity, 'temporary brittle bone disease' caused by deficiency of copper or vitamin C[,]. . . . [but h]e did not document how these diagnoses were made."[212] Also, Dr. Bays found that Dr. Paterson's symptomatology for TBBD looked suspiciously like the characteristics of child abuse. Specifically, Dr. Bays noted that "fractures, metaphyseal abnormalities, periosteal reaction, anterior rib changes, delayed bone age, vomiting and diarrhea,

---

[207] *See* Colin R. Paterson et al., *Osteogenesis Imperfecta: The Distinction from Child Abuse and the Recognition of a Variant Form*, 45 AM. J. MED. GENETICS 187 (1993) [hereinafter Paterson et al., *Distinction*]; Colin R. Paterson et al., *Osteogenesis Imperfecta Variant vs. Child Abuse: Reply*, 56 AM. J. MED. GENETICS 117 (1995) [hereinafter Paterson et al., *Reply*]; Colin R. Paterson et al., *Reply to Dr. Bawle: Temporary Brittle Bone Disease*, 49 AM. J. MED. GENETICS 132 (1994) [hereinafter Paterson et al., *Dr. Bawle*]; Paterson & McAllion, *supra* note 149, at 1451.

[208] Paterson & McAllion, *supra* note 149, at 1451–54.

[209] *Id.* at 1451.

[210] *Id.*

[211] *See* Bays, *supra* note 187, at 200; *see also* Paterson et al., *Distinction*, *supra* note 207, at 188.

[212] Bays, *supra* note 187, at 200.

apnea, hepatomegaly, anemia, and prematurity,"[213] which Dr. Paterson lists as symptoms of TBBD, are also all classic signs of child abuse and neglect.[214]

Controversy within the relevant scientific community may appear, at first, to be an obstacle to judicial comprehension. However, creative judges can use this information to their advantage by exploiting specific points of disagreement to expose flaws in an expert's theory, method, or conclusions. In the child abuse context, pediatric radiologists have expressed concerns about the scientific methodology employed in studies involving TBBD.[215]

Doctors Deborah Ablin and Shashikant Sane, writing in *Pediatric Radiology*,[216] explore how all medical experts use published scientific studies, when available, to bolster their trial testimony.[217] They suggest that courts should apply the *Daubert* criteria to identify various studies that: (1) were not subject to prepublication peer review[218] and (2) have no ascertainable error rates because "[n]o comprehensive detailed clinical information, detailed specific radiological findings of skeletal surveys, or other diagnostic imaging studies . . . are provided."[219] Doctors Ablin and Sane conclude that sometimes, as with TBBD, the medical literature is so incomplete and flawed that "objective analysis of the data by an independent observer is not possible."[220] Doctor Paul Kleinman has raised additional concern by observing that in studies conducted by Dr. Paterson and others ascribing injuries to TBBD and other variant forms of OI, "[m]ost of the radiologic features ascribed to transient brittle bone disease are those classically noted in cases of abuse."[221] According to Dr. Kleinman, Dr. Paterson's interpretations of radiographic images should be viewed with some skepticism. With respect to Dr. Paterson's 1993 article,[222] Dr. Kleinman has opined that "[b]ecause no radiologists were authors of this publication, and no details are given regarding the methods employed in the radiologic evaluation of these patients, it is difficult to assess the accuracy of these findings."[223]

---

[213] *Id.*

[214] *Id.*

[215] *See, e.g.*, *Re* X (Non-Accidental Injury: Expert Evidence), 2 F.L.R. 1, 18 (Royal Courts of Justice, Fam. Div. 2001) ("Dr. Paterson's two articles have not escaped criticism in the medical literature. They have however gathered but scant support.").

[216] *See* Ablin & Sane, *supra* note 24.

[217] *See id.* at 111–12.

[218] *Id.* at 111.

[219] *Id.* at 112.

[220] *Id.*

[221] Lachman et al., *supra* note 20, at 211.

[222] *See* Paterson et al., *Distinction*, *supra* note 207.

[223] Lachman et al., *supra* note 20, at 211.

3. *Applying the "Same Level of Intellectual Rigor" Standard*

In *Kumho Tire Co. v. Carmichael*,[224] the Supreme Court required that courts "make certain that an expert . . . employs in the courtroom the same level of intellectual rigor that characterizes the practice of an expert in the relevant field."[225] This was a warning to judges to scrutinize scientific evidence prepared specifically for litigation purposes. This is not a new concern. When *Daubert* was remanded, the Ninth Circuit specifically considered the question of whether the experts had "developed their opinions expressly for purposes of testifying."[226] Although it may be essentially impossible to discover science that is not tainted by an eye towards litigation in certain fields such as tobacco research, this factor alone is not an accepted basis for pretrial exclusion.[227]

Any judge assessing the intellectual rigor used by an expert to arrive at a diagnosis like TBBD might develop serious concerns. As early as 1995, TBBD was identified as a litigation-driven diagnosis.[228] That year the National Center for the Prosecution of Child Abuse ("NCPCA") issued a public warning that both OI and TBBD were becoming increasingly popular defenses in child abuse cases. In a one-page bulletin, the NCPCA described osteogenesis imperfecta as a "rare genetic disorder" and noted that medical experts place the range of OI births between one in 20,000 and one in 100,000. The NCPCA found general agreement in the medical literature that "the vast majority of OI cases are obvious and/or present no diagnostic difficulty if a thorough examination is conducted by a qualified physician."[229] After describing the four major types of OI, the NCPCA noted that the likelihood of OI presenting without typical symptoms in a way likely to be indistinguishable from child abuse is approximately 1 in 3,000,000.[230] The NCPCA also observed that "[s]tatistically, it makes no sense

---

[224] 526 U.S. 137 (1999).

[225] *Id.* at 152.

[226] Daubert v. Merrell Dow Pharms., Inc., 43 F.3d 1311, 1317 (9th Cir. 1995).

[227] *See* Lust *ex rel.* Lust v. Merrell Dow Pharms., Inc., 89 F.3d 594, 597 (9th Cir. 1996) (expressing concern that the expert was a "professional plaintiff's witness" and that his opinion might be influenced by a "litigation-driven financial incentive"); *cf.* Berry v. CSX Transp., Inc., 709 So. 2d 552, 569 (Fla. Dist. Ct. App. 1998) (describing how the court's decision to admit certain epidemiological studies was influenced by the fact that the research was "conducted independently of this litigation").

[228] *Questionable "Brittle Bone Disease" Defenses to Physical Abuse*, UPDATE (Am. Prosecutors Research Inst. Nat'l Ctr. Prosecution Child Abuse), Oct. 1995, at 1.

[229] *Id.*

[230] *Id.* This is supported by the relevant medical literature. *See, e.g.*, Jane M. Wynne & Christopher J. Hobbs, *Commentary, in* Roger Smith, *Osteogenesis Imperfecta, Non-Accidental Injury, and Temporary Brittle Bone Disease*, 72 ARCHIVES DISEASE CHILDHOOD 169, 172 (1995):

for the defense to claim that OI can easily be mistaken for child abuse. Its occurrence is rare while the occurrence of non-accidental injury in children is all too common."[231] In addition, when defense medical experts publicly express their concern about the "immense harm [that] can be done to families by the inaccurate diagnosis of non-accidental injury,"[232] they emphasize only one side of the problem and omit the dangers that arise when abuse is not reported or when inaccurate diagnoses of accident are accepted by judges and juries.

B. *Second: How Did the Experts Test Their Conclusions?*

To learn how judges should explore scientific evidence and assess an expert's conclusions, a child abuse case from New York provides a helpful model because it details a judge's effort to assess whether initial diagnoses of child abuse had been subjected to valid testing methods. This case involved a two-month-old child who suffered at least 17 fractured bones while in his parents' care.[233] Judge Nora Freeman of the Queens County Family Court of New York[234] presided over extensive pretrial hearings to determine the admissibility of proffered defense expert testimony attributing the infant's injuries to a mild form of Osteogenesis Imperfecta. This was the only explanation offered by defendant parents to explain their son's injuries.

After a five-day hearing, Judge Freeman made detailed findings of fact that demonstrate a clear understanding of the complex medical and legal questions presented to the court. Judge Freeman began by explaining what she had learned about the prevalence rate of OI and the severity of this disease in most of its forms. According to Judge Freeman:

> Osteogenesis imperfecta is described by the three physicians who testified at trial as an extremely rare condition, observed in approximately one birth per 250,000. The bones of a newborn afflicted with the most severe form of OI will fracture during the birth process and also during routine handling. Such a baby is unlikely

---

> [T]he probability that an individual infant with no relevant family history has osteogenesis imperfecta where the skeleton is normal, there are no wormian bones, there is no or trivial history of trauma, the infant is not weight bearing yet has a fractured skull, ribs, or metaphyseal fractures is in the Taitz range of probabilities, that is, [in the] millions.

Id.

[231] *Id.*

[232] Colin Paterson, *The Child With Unexplained Fractures*, 147 NEW L.J. 648, 648 (1997).

[233] *In re* Matthew D., 641 N.Y.S.2d 526 (1996).

[234] *See supra* note 42 (explaining that New York has retained the *Frye* standard).

> to survive infancy. Milder forms of OI result in repeated fractures which may be reduced by careful training for the caretakers.[235]

With this knowledge of OI's prevalence, Judge Freeman next described how understanding the diagnostic process helps a court assess the validity of the scientific methodology.

> Diagnosis of OI is based on several factors, including genetic history (parents and siblings); the type of fractures (typically, the long bones are fractured in more than one site); presence of "Wormian bones" in the skull (irregularities in the frontal sutures, visible in x-rays); blue or bluish sclerae; and a triangular shape to the face. In addition to clinical observations, OI can also be confirmed by various blood tests for the child and parents. The most sophisticated test, performed only rarely, requires a biopsy from which unusual levels of collagen can be detected. . . . The Seattle biopsy is recognized as conclusive in 85% of cases, meaning that 15% of such test results will be negative *despite* the patient actually having OI.[236]

During a lengthy pretrial hearing, the state offered two expert witnesses: (1) an expert in pediatrics and child abuse and (2) an expert in pediatric radiology.[237] After listening to these experts, Judge Freeman concluded that the validity of any diagnosis must account for the fact that certain types of fractures are typical of child abuse. According to Judge Freeman, the government experts "established that . . . several [fractures], described as 'bucket-handle fractions [sic],' were typical of child abuse."[238] The court also understood and considered the medical significance of the fact that all clinical findings and test results in this case were negative for OI. The findings revealed that "blood tests to detect genetic abnormalities were negative; that there was no known family history of OI; that there was no sign of 'Wormian bones' characteristic of OI; and that the baby suffered no further fractures during his hospitalization."[239]

Based on these findings, both of the experts who testified for the state unequivocally ruled out OI as a diagnosis.[240] During the hearing, these experts testified that they had

> concluded to a reasonable degree of medical certainty that the injuries to the baby were caused by trauma inflicted by an adult. Both doctors also testified that they learned subsequently that the baby had not suffered any new fractures after his discharge from the hospital to a series of foster homes, and that the Seattle

---

235 *In re Matthew D.*, 641 N.Y.S.2d at 528.

236 *Id.* at 528.

237 *Id.* at 528–30.

238 *Id.* at 528–29.

239 *Id.* at 529.

240 *See id.*

> biopsy was negative. Each doctor testified that those two additional factors supported and strengthened their conclusions, reached months earlier, that there was no medical support for a diagnosis of OI, "none whatsoever."[241]

After careful analysis of the government's medical expert evidence focused on their diagnostic methodologies, the court was satisfied that the experts had reliably tested their conclusions.

Judge Freeman then assessed the validity of the defense expert's testimony. The defense expert witness, a pediatric orthopedist with fifteen years of experience as Director of the OI Clinic at the Hospital for Special Surgery ("HSS"),[242] testified that "in his opinion Lucas suffered from a 'mild' form of OI."[243]

Judge Freeman noted that, despite the defense expert's credentials, his testimony was undermined by the unreliable methods that he used to test his initial diagnosis.[244] First, Judge Freeman specifically identified the following flaws in the defense expert's methodology: (1) he was not sure whether he had reviewed all of the baby's medical records; (2) he testified that the baby had not suffered a skull fracture when the x-rays showed a skull fracture; and (3) he relied solely on the parents and their attorney for his information.[245] Second, the court noted that the defense expert's diagnosis of "mild OI," with a period of remission coinciding with the baby's foster care placement did not account for the facts that the Seattle biopsy OI test result was negative or that the baby sustained no new fractures during the ten months that he had been removed from the defendants' care.[246] Judge Freeman specifically questioned the basis of the defense expert's spontaneous remission theory, pointing out that the only support for spontaneous remission was the expert's claim that he had seen it happen before.[247]

According to Judge Freeman, the only point of agreement among the three experts was their shared opinion that bucket handle fractures, of the type sustained by the infant in this case, are highly specific to child abuse.[248] Ultimately, Judge Freeman was unpersuaded by the defense expert's subsequent

---

[241] *In re Matthew D.*, 641 N.Y.S.2d at 529.

[242] *Id.*

[243] *Id.*

[244] *Id.*

[245] *Id.*

[246] *Id.* at 530.

[247] *In re Matthew D.*, 641 N.Y.S.2d at 529–30. It should be noted that Judge Freeman's skepticism on this point is consistent with the medical literature, which indicates that "the child who has multiple unexplained fractures in one environment and then has no further fractures when removed from that environment should be suspected of having nonaccidental trauma." Bays, *supra* note 187, at 201 (quoting S. Gahagan & M.E. Rimsza, *Child Abuse or Osteogenesis Imperfecta: How Can We Tell?*, 88 PEDIATRICS 987 (1991)).

[248] *In re Matthew D.*, 641 N.Y.S.2d at 530.

testimony explaining that a child with OI might sustain bucket handle fractures by falling from a bicycle.[249] Her conclusion was based on a careful analysis of the complex medical evidence that both sides had presented. Judge Freeman demonstrated that she had developed a working understanding of base rate comparisons, the diagnostic criteria for OI, testing protocols, fracture etiology, and spontaneous remission. This case might serve as a model providing a detailed in-context look at careful judicial review of conflicting evidence.

C. *Third: How Did the Experts Rule Out Other Conclusions?*

The Advisory Committee Notes to Federal Rule of Evidence 702 specifically instruct judges to explore "whether the expert has adequately accounted for obvious alternative explanations."[250] If expert testimony diagnosing OI, a variant form of OI, or TBBD is admitted at trial, the defense attorney can argue that fractures happen more easily by accident or without the necessary level of intent to establish culpability. All of the physical evidence, however, must be viewed together and "[i]f the child has other clinical manifestations of physical abuse, such as bruises not associated with the site of a fracture, intracranial injuries, or retinal hemorrhages, it is extremely unlikely that the fractures are due to OI."[251] When the physical evidence is less definitive, Rule 702 suggests that judges should assess how reliably the defense expert has ruled out the only other possible cause—child abuse.

1. *Child Abuse and Osteogenesis Imperfecta*

a. *The Differential Diagnosis: Child Abuse v. Osteogenesis Imperfecta*

In 1997, two pediatric radiologists published an article comparing the indicia of TBBD that had been described in three articles appearing in the medical literature between 1993 and 1995 to child abuse.[252] These authors noted the following similarities between the classic indications of child abuse and the features cited as indicia of TBBD.[253]

---

[249] *Id.* Judge Freeman's skepticism may be attributable to the fact that the infant victim was only nine weeks old when his two dozen fractures were discovered. *Id.*

[250] FED. R. EVID. 702 advisory committee's note (citing Claar v. Burlington N.R.R., 29 F.3d 499 (9th Cir. 1994)).

[251] Gahagan & Rimsza, *supra* note 112, at 991.

[252] *See* Chapman & Hall, *supra* note 131. The three articles described were: Paterson et al., *Reply*, *supra* note 207, at 117–18; Paterson et al., *Dr. Bawle*, *supra* note 207, at 132; and Paterson et al., *Distinction*, *supra* note 207, at 187–92.

[253] Chapman & Hall, *supra* note 131, at 108.

Features of TBBD	Features of Child Abuse
Fractures during the first year of life.	The first year is the peak of fractures due to child abuse.
Fractures found "by accident" when radiographic images taken.	This is the typical mechanism for discovering fractures associated with child abuse because some fractures become evident only on follow-up studies.
Fractures occurring in the hospital.	This is where most fractures become evident and abuse can occur anywhere.
A high incidence of vomiting and diarrhea.	This is common in infancy and may be a presenting symptom of child abuse.

b. *Ruling Out Child Abuse: The Absence of Bruising*

The absence of bruising in children with fractures has been cited as compelling medical evidence that the injury was not caused by abuse. For example, Dr. Paterson has noted that, in patients with OI, "fractures may be accompanied by less superficial evidence of injury than would be expected if the bones had been normal."[254] This conclusion is based on the assumption that "[t]he force necessary to fracture a normal bone is thought to result invariably in external evidence of trauma"[255] (e.g., bruising) and that "the force required to fracture the bone was minimal, which implies weakness of the underlying bone—perhaps due to a temporary abnormality such as copper deficiency or subtle forms of osteogenesis imperfecta."[256]

More recent medical literature indicates that this conclusion should be questioned when it forms part of the methodology of diagnosing either OI, a variant form of OI, or TBBD. For example, in a 1998 study, Drs. Matthew, Ramamohan, and Bennet found that in "normal children most fractures (91%) were not associated with bruising at the time of presentation."[257] Instead, this study showed that 72% of the fractures remained without evident bruising in the first week after injury. These data led the researchers to conclude that "the absence of bruising cannot be taken to imply either underlying bone disease or an increased possibility of non-accidental injury."[258] Doctor Kleinman has also

---

[254] Paterson et al., *Distinction*, *supra* note 207, at 188.

[255] M.O. Matthew et al., *Importance of Bruising Associated with Paediatric Fractures: Prospective Observational Study*, 317 BRIT. MED. J. 1117, 1117 (1998).

[256] *Id.* at 1118.

[257] *Id.*

[258] *Id.*

noted that "the vast body of child abuse literature . . . indicates that bruises and other signs of trauma are frequently absent in abused infants."[259]

Orthopedic doctors have assessed the validity of the methodology used by some physicians to support the conclusion that a lack of bruising indicates OI.

> Many doctors are now involved in the care of children with fractures, particularly in cases where child abuse is suspected. Some have assumed that the lack of bruising means that a pathological process such as osteogenesis imperfecta is present and that the bone has fractured easily without the use of undue force and therefore is not a non-accidental injury. The work on which these ideas are based has tended to appear in the letters section rather than the peer reviewed sections of medical journals. In suspected child abuse, however, the fact that breaks and bruises do not always occur together can have more serious consequences.[260]

These findings indicate that expert testimony indicating that child abuse was ruled out, in part based on the lack of bruising, should be carefully scrutinized by the court.

2. *Ruling Out Child Abuse: Abuse Injuries Compared to Accidental Injuries*

There are medical studies that compare the type of fractures typically caused by child abuse to those typically caused by accidental injury. In a recent study, Drs. Paula Schweich and Gary Fleisher reviewed the medical charts of twenty-one children hospitalized for rib fractures. Most of the fractures (76%) had been caused by accidents, while 24% were attributed to abuse. Children with accidental injuries were significantly older than children subjected to abuse.[261] The mean age of children injured in accidents was eight-and-seven-twelfths, and they ranged from two to fifteen years old.[262] The mean age of children injured through abuse was three months, and they ranged from two weeks to seven months old.[263] These findings make sense because younger children are more often victimized while older children are more likely to engage in activities that result in accidental fractures. It also makes sense that the victims of accidental injury had clinical histories of sudden forceful trauma, for example, motor vehicle accidents, falls from heights, and gunshots.[264] By contrast, victims of abuse commonly presented

---

[259] Lachman et al., *supra* note 20, at 211.

[260] Deborah Eastwood, *Breaks Without Bruises Are Common and Can't Be Said to Rule Out Non-Accidental Injury*, 317 BRIT. MED. J. 1095, 1095 (1998) (citations omitted).

[261] Paula Schweich & Gary Fleisher, *Rib Fractures in Children*, 1 PEDIATRIC EMERGENCY CARE 187, 188 (1985).

[262] *Id.*

[263] *Id.*

[264] *Id.*

with unexplained respiratory distress.[265] "The accidental group [also] had fewer rib fractures (average, 3.3 fractures; range, one to eight) than the abused group (average, 11.8 fractures; range, three to 23)."[266] These data provide additional criteria to consider when experts purport to differentiate between injuries caused by accident and injuries caused by abuse.

VII. CONCLUSION

> Expert witnesses may wish to win their case, but this should not be done at the expense of the facts and is not necessarily in the best interests of the child. . . . In an ideal world the view given by the expert should be straightforward, not misleading or biased and well researched; a well balanced and non-partisan view will be more welcome to the court than fixed ideas and an inability to consider all sides of the problem.
>
> Dr. Roger Smith[267]

> Trial judges must be on guard against all forms of junk science that may creep into the courtroom.
>
> *Greenwell v. Boatwright*[268]

Legal accuracy in cases where science and law intersect will only improve if judges and experts reevaluate their respective roles. If, after a decade of *Daubert*, judges still struggle with basic scientific concepts, the reevaluation process requires a context-specific approach to specific science-law questions. An effective program should also shun any approach to science-law questions that ignores the dynamic nature of adjudication and the impact of the trial process on how questions are asked and answered. For example:

> we should also avoid "explaining away," as an *a priori* epistemological problem, the presence of scientific disagreement in legal contexts. In some instances, a legal setting may be drawing on pre-existing scientific disagreement, yet, in others there may be special features of the legal setting itself which are contributing to the disagreement in question. Scientific disagreements in legal settings should be empirically investigated with consideration of the particulars of the scientific knowledge claims in question, the specific features of the legal setting in question, and the specific way science and law have been brought together.[269]

---

[265] *Id.*

[266] Bays, *supra* note 187, at 197.

[267] Smith, *supra* note 230, at 171.

[268] 184 F.3d 492, 501 (6th Cir. 1999).

[269] Edmond & Mercer, *supra* note 14, at 19.

This should help us to recognize "[w]hat is frequently not appreciated by both professionals in the field and by the public, [which] is the extent to which the legal system facilitates irresponsible expert testimony."[270]

We have seen how child abuse cases often rely on the testimony of medical experts. Because all parties have access to medical experts, judges must determine whether and to what extent doctors may testify given the potential influence of medical expert evidence on the legal outcome.[271]

At least one medical commentator has opined that experts may be partly to blame for the problem and partly the source of any future cure. "[P]hysicians have been quick to condemn the legal profession as the cause for the surge in [for example,] medical malpractice lawsuits. However, in reality, the greater impetus has been the medical expert witness who has developed unique theories of causation with consequent corruption of science."[272] Courts cognizant of the influence of expert evidence on jury decisions should scrutinize all medical opinions, and, if the underlying evidence is admitted, continue to carefully monitor experts, so that they cannot draw unwarranted inferences.

> Although the offering of an opinion is a process that is somewhat less objective than the observation of a fact, expert witnesses are still sworn to tell the truth. A reasonable expectation about the meaning of truth telling in the context of offering of an opinion might be that if an opinion is divergent from generally accepted medical knowledge, the expert should acknowledge this fact and should avoid "unique theories of causation" and other irresponsible positions. For example, if the expert must hypothesize a previously undescribed medical condition to explain pathology, the conjectural nature of this stretch should be pointed out as part of the duty to tell "the whole truth."[273]

Finally, judges should be wary of experts who may be unqualified to testify to a particular opinion. This can be the result of a lack of experience (e.g., residency, fellowship, clinical practice, clinical research in a particular area of medicine), a

---

[270] Holmgren, *supra* note 127, at 1.

[271] Some medical experts openly acknowledge the link between their effectiveness as an expert witness and their professional status. Dr. Paterson, has been quoted as saying: "I am extremely anxious not to make a mistake of diagnosing brittle bone disease when the reality is child abuse. The effect of that on my reputation, apart from anything else would be quite devastating." *Doctor's Doubts on Abuse Cases*, HERALD (Glasgow), Jan. 21, 1997, at 9, LEXIS, Nexis Library, GHERLD File.

[272] Michael I. Weintraub, *Expert Witness Testimony: A Time for Self-Regulation?*, 45 NEUROLOGY 855, 855 (1995) (suggesting that medical experts should be subjected to careful peer review and that experts who testify irresponsibly should be exposed).

[273] David L. Chadwick & Henry F. Krous, *Irresponsible Testimony by Medical Experts in Cases Involving the Physical Abuse and Neglect of Children*, 2 CHILD MALTREATMENT 313, 314 (1997).

lack of formal qualifications (e.g., board certification or advanced specialized certification), or both.

The scientific literature includes suggestions that medical experts be required to demonstrate relevant training and experience in cases similar to the case in which they have been called on to testify.[274] A more specific and more recent physician proposal would require that medical experts in child abuse cases "document for the [c]ourt the following:"

> 1. General training or experience in child abuse and neglect.
> 2. Specific training or experience relative to the particular type of case being adjudicated.
> 3. Memberships in relevant professional societies.
> 4. Child abuse and neglect conference presentations and attendance.
> 5. Relevant professional publications.[275]

Applying this standard, clinicians, diagnosticians (e.g., radiologists), and pathologists would be required to demonstrate that they "have knowledge of natural, medical disorders associated with bone fractures, easy bruising, and ostensible sudden, unexpected infant death."[276] Courts would learn that in many cases the "testimony of pediatric radiologists can be crucial in the differentiation of child abuse from accidental fractures, osteopenia of prematurity, and bone disorders, such as osteogenesis imperfecta."[277]

There is only one way to prevent a misunderstanding of science from creating a miscarriage of justice. When legal accuracy depends on scientific validity, judges and experts should share a common goal: "[T]he time has come for physicians and lawyers and their respective professional societies to begin a process by which such unsavory testimony can be exposed, peer reviewed, and ultimately prevented."[278] This process requires that judges confront their ignorance, learn from their mistakes, and think creatively about how to better understand and use scientific information. If judges do not slam the gate on questionable scientific evidence, we can have little confidence in our legal determinations. It takes time and energy to comprehend the science behind any causal theory or medical diagnosis. As society becomes more scientifically complex, judges will need to spend more time, in the long term, honing their scientific skills and, in the short term, developing questions and protocols that make their inquiries more logical and comprehensible. Child abuse cases illustrate

---

[274] *See* David L. Chadwick, *Preparation for Court Testimony in Child Abuse Cases*, 37 PEDIATRIC CLINICS N. AM., 955, 958 (1990).

[275] Chadwick & Krous, *supra* note 273, at 320.

[276] *Id.*

[277] *Id.*

[278] *Id.*

why an accurate understanding of science is vital; because the life of a child—and not some abstraction of justice —hangs in the balance.[279]

[279] According to Dr. Christine Hall, consultant pediatric radiologist at the Great Ormond Street Children's Hospital in London:

> The hypothetical condition temporary brittle bone disease bears a striking similarity to many cases of non-accidental injury. I would suggest that they are the same condition but with different labels depending on the credibility of the child caretaker's explanation. I know of one case, where Dr. Paterson's theory was accepted, the baby was taken off the "at risk" register and returned home, and subsequently died . . . .

Annabel Ferriman, *Accused of Child Abuse. But Was His Baby a Victim of Brittle Bone Disease?*, INDEPENDENT (London), March 18, 1997, at 3, LEXIS, Nexis Library, INDPNT File (quotation marks omitted).