

7-1-1999

A Ray of Light For Judges Blinded by Science: Triers of Science and Intellectual Due Process

Erica Beecher-Monas

University of Arkansas at Little Rock, e.beecher@wayne.edu

Recommended Citation

Erica Beecher-Monas, *A Ray of Light For Judges Blinded by Science: Triers of Science and Intellectual Due Process*, 33 Ga. L. Rev. 1047 (1999).

Available at: <https://digitalcommons.wayne.edu/lawfrp/201>

This Article is brought to you for free and open access by the Law School at DigitalCommons@WayneState. It has been accepted for inclusion in Law Faculty Research Publications by an authorized administrator of DigitalCommons@WayneState.

A RAY OF LIGHT FOR JUDGES BLINDED BY SCIENCE: TRIERS OF SCIENCE AND INTELLECTUAL DUE PROCESS

Erica Beecher-Monas*

I. INTRODUCTION

Judges, traditionally triers of law, occasionally pressed into service as triers of fact, now must also be triers of science in cases where experts proffer scientific evidence.¹ Years after *Daubert v. Merrell Dow Pharmaceuticals, Inc.*² made judges responsible for assessing scientific validity, judges are still grappling with the fact that they can no longer merely count scientific noses³ but must instead decide whether expert testimony meets the criteria of good science.⁴ Predictably, not everyone is pleased with this new state of affairs, and many question judicial competence in this area.⁵ But

* Assistant Professor, University of Arkansas at Little Rock School of Law. LL.M. Columbia University School of Law, M.S. (Anatomy) University of Miami. This article is written in partial fulfillment of the J.S.D. requirements of the Columbia University School of Law. The author is deeply grateful to the late Curtis J. Berger for his encouragement. Many thanks also go to Frank P. Grad, Martha A. Fineman and to Theresa Beiner for their helpful comments and criticism. For her unflagging enthusiasm, insight, and very able research assistance, the author thanks Julie Peters Zamacona.

¹ *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579 (1993).

² *Id.*

³ Before *Daubert*, the federal courts overwhelmingly applied a consensus standard for admissibility. This was the standard of *Frye v. United States*, 293 F. 1013, 1014 (D.C. Cir. 1923), which explained that scientific testimony must "be sufficiently established to have gained general acceptance in the particular field in which it belongs."

⁴ *Daubert*, 509 U.S. at 592-93.

⁵ Skeptical about the feasibility of the Supreme Court's mandate, scholars, practitioners, and journalists have all expressed reservations about the abilities of judges to assess questions of scientific validity. See Daniel J. Capra, *The Daubert Puzzle*, 32 GA. L. REV. 699, 703 (1998) (arguing that scientists can assess reliability better than judges); Adina Schwartz, *A "Dogma of Empiricism" Revisited: Daubert v. Merrell Dow Pharmaceuticals, Inc. and the Need to Resurrect the Philosophical Insight of Frye v. United States*, 10 HARV. J.L. & TECH. 149, 153 (1997) (castigating *Daubert's* "intellectual bankruptcy"); Jennifer Laser, Note, *Inconsistent Gatekeeping in Federal Courts: Application of Daubert v. Merrell Dow Pharmaceuticals, Inc. to Nonscientific Expert Testimony*, 30 LOY. L.A. L. REV. 1379, 1379 (1997) (arguing that increasing reliance on expert testimony has resulted in inconsistent judicial decisionmaking); *Developments in the Law—Confronting the New Challenges of Scientific Evidence*, 108 HARV. L. REV. 1481, 1493-1509 (1995) (arguing that the *Daubert*

judges are not unarmed for these decisions.

The Supreme Court gave judges some rudimentary guidelines in *Daubert*, outlining the notions of scientific validity and fit.⁶ The Supreme Court reiterated and amplified these concepts in *General Electric Co. v. Joiner*.⁷ In addition, the Federal Judicial Center has published a reference manual for evaluating scientific evidence which explains key aspects of scientific disciplines likely to be the subject of expert testimony.⁸ Courses have sprung up to help familiarize judges with scientific issues.⁹ And, if all else fails, the trial court may appoint its own experts for advice.¹⁰ One analytic framework that is generally overlooked, however, is the guidance of federal regulatory agencies, which have been making sound scientific validity determinations for generations.

analysis is essentially indeterminate and cannot account for results in particular cases); see also Arthur Kantrowitz, *Proposal for an Institution for Scientific Judgment*, 156 SCIENCE 763, 764 (1967) (proposing science court to solve problem of judicial inability to handle expert testimony); John L. Thornton, *Courts of Law v. Courts of Science: A Forensic Scientist's Reaction to Daubert*, 1 SHEPHARD'S EXPERT AND SCI. EVIDENCE Q. 475, 482 (1994) (noting that although much of forensic science has "precious little scientific foundation" judges continue to find it admissible post-*Daubert*); cf. *McKnight v. Johnson Controls, Inc.*, 36 F.3d 1396, 1406 (8th Cir. 1994) (describing plaintiff's argument that *Daubert* makes expert testimony more readily admissible, while defendant argued that *Daubert* makes expert testimony less readily admissible). See generally PETER W. HUBER, GALILEO'S REVENGE: JUNK SCIENCE IN THE COURTROOM (1991). Judges themselves have expressed doubts about their capabilities. Marcia Coyle, *Cert. Granted For Expert Witness Case*, NAT'L L.J., Mar. 31, 1997, at B1. Chief Justice Rehnquist was prominent among the judges complaining about the gatekeeping role assigned to judges by the *Daubert* majority. See *Daubert*, 509 U.S. at 599-600 (Rehnquist, C.J., concurring in part and dissenting in part). Judge Kozinski, on remand, was more vociferous, calling the task of evaluating expert testimony "a far more complex and daunting task in a post-*Daubert* world than before." *Daubert v. Merrell Dow Pharms., Inc.*, 43 F.3d 1311, 1315-16 (9th Cir. 1995).

⁶ *Daubert*, 509 U.S. at 593-94.

⁷ 522 U.S. 136 (1997). *Joiner* appears to be far from the last word. The Supreme Court recently considered *Daubert* issues in the context of non-scientific expert testimony in *Kumho Tire Co. v. Carmichael*, 67 U.S.L.W. 4179 (U.S. Mar. 30, 1999).

⁸ FEDERAL JUDICIAL CTR., REFERENCE MANUAL ON SCIENTIFIC EVIDENCE (1994).

⁹ See John M. Conley & David W. Peterson, *The Science of Gatekeeping: The Federal Judicial Center's New Reference Manual on Scientific Evidence*, 74 N.C. L. REV. 1183, 1204 (1996) (referring to the authors' courses for judges.)

¹⁰ See FED. R. EVID. 706. Federal Rule of Evidence 706 provides for court-appointed experts, and *Daubert* suggested that judges should take advantage of this rule. Some judges do appoint their own experts. See, e.g., *Hall v. Baxter Healthcare Corp.*, 947 F. Supp. 1387, 1392 (D. Or. 1996) (describing basis for appointing independent advisors to court). This option is not without its drawbacks, however. The court in *Hall* misunderstood and mischaracterized its own experts, choosing some of their conclusions over others without explanation. See *id.*

Lessons from the federal regulatory agencies are particularly salient because like judges, most agency decisionmakers are not trained scientists, yet they must make creditable scientific validity assessments. Through a process of articulating assessment guidelines, soliciting comment from scientists and other interested parties on the assessment guidelines, and requiring agency decisionmakers to explain fully their determinations on the basis of an adequate record subject to judicial review,¹¹ the agencies have developed a standard of intellectual due process which has much to offer federal judges.

In particular, the Environmental Protection Agency has developed a sensible set of Proposed Guidelines reflecting not only the current scientific understanding of how such evidence about toxic substances should be assessed but also incorporating explicit policy considerations and explaining the basis and rationale for each. While there are important limitations to a wholesale adoption of agency decisionmaking criteria, the approach is well-grounded in science. A similar—though not identical—approach, in conjunction with the guidance of the Supreme Court guidelines and the Federal Judicial Center's Manual, would enable federal trial judges to make more intellectually defensible admissibility determinations based on science, logic, and law.

This Article discusses and critiques the analytic frameworks available to judges for their admissibility determinations. Focusing on the troubling scientific admissibility issues presented in toxic tort litigation, Part II of this Article discusses the failures of the Supreme Court's *Daubert* and *Joiner* cases to provide the guidance judges need to make scientifically adequate admissibility determinations and illustrates these failings through a critique of their toxic tort progeny. Part III appraises the guidance given in the Federal Judicial Center's *Reference Manual on Scientific Evidence*. That Part explains why, though the Reference Manual is helpful in understanding the principles and methodology underlying several

¹¹ See Proposed Guidelines for Carcinogen Risk Assessment, 61 Fed. Reg. 17,960, 17,960 (1996) [hereinafter Proposed Guidelines] (providing an example of administrative regulation process). Notice and comment rulemaking procedures are mandated by the Administrative Procedure Act, 5 U.S.C. §§ 551, 553 (1994).

scientific disciplines, it is not enough. Judges—and the lawyers responsible for educating them about their cases—need to know more than what goes into optimal experimental design. What judges and lawyers need to know is not how to design the best scientific study, but how to assess imperfect ones.¹²

Assessing imperfect studies—that is, the scientific validity of conclusions drawn from imperfect knowledge—is the explicit mandate of the Environmental Protection Agency's (EPA) Proposed Guidelines.¹³ Part IV summarizes the guidelines proposed by the EPA to assess scientific validity and compares the composite requirements of *Daubert* and *Joiner* with the requirements of administrative agencies such as the EPA. This Part analyzes the applicability to judicial determinations of agency guidelines aimed at protecting the public health, the object of which is to assess liability. Part IV argues that while the EPA guidelines cannot be adopted in whole cloth, they offer valuable insights that may make judicial decisionmaking more intellectually defensible and fair to the litigants.

This Article concludes that although separately each set of guidelines available to federal judges is insufficient to make sound scientific validity determinations, together they illuminate important facets of the determination. These guidelines can be combined with an understanding of the process of science and the nature of probabilistic reasoning to form a sound framework for analysis. The purpose of this Article is to offer some suggestions for that process.

II. THE SUPREME COURT'S CRITERIA

The federal trial judge faced with a proffer of scientific testimony must first determine whether the expert's testimony is "scientific knowledge" that will assist the factfinder to determine a fact in

¹² No study is ever perfect, and those relied on by litigants are often less perfect than most. See Jack L. Landau & W. Hugh O'Riordan, *Of Mice and Men: The Admissibility of Animal Studies to Prove Causation in Toxic Tort Litigation*, 25 IDAHO L. REV. 521, 550 (1989). One of the key problems is that animal toxicity studies are not designed to demonstrate causation but to identify biological mechanisms of toxicity. See *id.*

¹³ See Proposed Guidelines, *supra* note 11 (detailing carcinogen assessment and reassessment procedures).

issue.¹⁴ To qualify as scientific knowledge, the Supreme Court held, “an inference or assertion must be derived by the scientific method.”¹⁵ Assisting the factfinder, the Court explained, is the issue of relevance, or “fit.”¹⁶ In *Daubert*, the Supreme Court’s explanation of both these precepts—scientific method and “fit”—was cursory at best. *Joiner*¹⁷ expanded the discussion of “fit” somewhat, but still left many questions unanswered.

A. DAUBERT’S EXPLANATION OF SCIENTIFIC METHOD

Scientific validity and “good grounds” are synonymous, according to the Court, and for scientific evidence good grounds implicates the scientific method.¹⁸ In order to decide whether an inference or assertion was derived by the scientific method and thus constitutes scientific knowledge, *Daubert* instructs the trial judge to assess the scientific validity of the reasoning underlying the testimony.¹⁹ This assessment must be performed whenever scientific evidence is contested.²⁰ Anything purporting to be scientific knowledge, short

¹⁴ See *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 590 (1993) (stating “[t]he subject of an experts’ testimony must be ‘scientific . . . knowledge’”).

¹⁵ *Id.* at 589.

¹⁶ *Id.* at 591.

¹⁷ *Joiner* involved claims that the plaintiff’s lung cancer had been promoted by his exposure to the defendants’ PCBs, which the plaintiff sought to support by expert testimony founded on two animal studies and four epidemiology studies. *General Elec. Co. v. Joiner*, 522 U.S. 136 (1997). The federal trial judge’s admissibility determination starts with a two-pronged inquiry into “whether the reasoning or methodology underlying testimony is scientifically valid and . . . whether that reasoning or methodology properly can be applied to the facts in issue.” *Daubert*, 509 U.S. at 592-93.

¹⁸ 509 U.S. at 590-91.

¹⁹ *Id.* at 593.

²⁰ Unlike *Frye*, the Court did not limit its requirements to novel scientific evidence. The decision in *Frye*, in distinction, was expressly limited to novel scientific evidence. See *Frye v. United States*, 293 F. 1013, 1014 (D.C. Cir. 1923) (explaining exclusion of expert testimony which is not “sufficiently established to have gained general acceptance”). Courts frequently get this distinction confused and misread *Daubert* as also limited to novel scientific evidence. See, e.g., *Mitchell v. United States*, 141 F.3d 8 (1st Cir. 1998) (finding *Daubert* to apply only to novel scientific evidence); *Masayeva v. Hall*, 118 F.3d 1371 (9th Cir. 1998) (same); *Thornton v. Caterpillar, Inc.*, 951 F. Supp. 576, 577 (D.S.C. 1997) (same); *Lappe v. American Honda Motor Co.*, 857 F. Supp. 222, 228 (N.D.N.Y. 1994) (same). This reading is wholly unwarranted.

of scientific law "such as the laws of thermodynamics," must be subjected to analysis.²¹

The Court's sole explanation of scientific method was that scientific knowledge "implies a grounding in the methods and procedures of science."²² Although the Court did not offer a description of these methods and procedures, it did outline four "general observations" to guide the inquiry.²³ They are testability, peer review and publication, error rate, and general acceptance. These "flexible guidelines"²⁴ incorporate not only the Supreme Court's notion of the scientific method but also its assessment of the importance of peer review to scientists.

1. *Testability.* Testability, as the Supreme Court emphasized, is the cornerstone of the scientific method.²⁵ The importance the Supreme Court gave to the concept of testability reflects its understanding that credibility in the scientific community is expressed in terms of a scientist's adherence to these standards.²⁶ Within this concept, the Supreme Court included ideas about implementing standards for empirical investigation and testing falsifiable hypotheses.²⁷ Citing Karl Popper, one of the preeminent philosophers of science, the Supreme Court explained that falsifiability was what distinguishes science from nonscience.²⁸

²¹ *Daubert*, 509 U.S. at 592 n.11.

²² *Id.* at 590.

²³ *See id.* at 594 (characterizing inquiry as a "flexible one").

²⁴ *Id.*

²⁵ *See id.* at 593 (explaining that "[s]cientific methodology today is based on generating hypotheses and testing them to see if they can be falsified") (citing Karl Popper, *CONJECTURES AND REFUTATIONS: THE GROWTH OF SCIENTIFIC KNOWLEDGE* 37 (5th ed. 1989)).

²⁶ *Daubert*, 509 U.S. at 593. Falsifiability is the term defining the scientific method that was picked up by the Supreme Court in explaining the concept of testability, much to the chagrin of Chief Justice Rehnquist, who in the *Daubert* dissent remarked that he was "at a loss to know what is meant when it is said that the scientific status of a theory depends on its 'falsifiability' . . ." *Id.* at 600 (Rehnquist, C.J., concurring in part and dissenting in part). According to Karl Popper, the distinguishing characteristic of a scientific statement is "its falsifiability, or refutability, or testability." *Id.* at 593.

²⁷ *Id.* at 593. As one scientist described the importance of testability: "The goal of science is the systematic organization of knowledge about the universe on the basis of explanatory principles that are genuinely testable." Francisco J. Ayala, *Biology as an Autonomous Science*, 56 AM. SCI. 207, 207 (1968).

²⁸ *Daubert*, 509 U.S. at 593. Popper argued that science was distinguishable from other disciplines—specifically psychoanalysis and economics—because it is possible to devise a test which would disprove a scientific hypothesis but not an economic one. *See* KARL R. POPPER,

Other than to state the importance of the concept, however, the Court was silent on its meaning.²⁹ The Court neither explained nor gave examples of the testability concept's application.

2. *Peer Review and Publication* and 3. *General Consensus*. The Supreme Court recognized the importance of feedback to scientists.³⁰ The Court explained that submitting an expert's work to the scientific community is "a relevant, though not dispositive, consideration in assessing the scientific validity of a particular technique or methodology on which an opinion is premised."³¹ Unfortunately, however, two of the Court's four factors are devoted to social feedback mechanisms, giving them unwarranted importance and risking a return to the old *Frye* standard despite having expressly overruled it.³² This overemphasis is problematic because although

THE LOGIC OF SCIENTIFIC DISCOVERY 86 (1959); KARL R. POPPER, CONJECTURES AND REFUTATIONS: THE GROWTH OF SCIENTIFIC KNOWLEDGE 36-37 (5th ed. 1989). Other philosophers of science have disputed this basic assumption, claiming that science, like other modes of human thought, is fundamentally indeterminate. See, e.g., THOMAS S. KUHN, THE STRUCTURE OF SCIENTIFIC REVOLUTIONS (1962) (describing history of scientific method). Even absent the concept of an objectively determinate "scientific method," however, "commitment in science reaches beyond argument and rests ultimately on agreed-upon procedures." ALAN G. GROSS, THE RHETORIC OF SCIENCE 32 (1990).

²⁹ *Daubert*, 509 U.S. at 593. The Supreme Court recognized that "a key question in determining whether a theory or technique is scientific knowledge that will assist the trier of fact will be whether it can be (and has been) tested." *Id.* Other than a citation to Karl Popper, linking the definition to "falsifiability or refutability," however, the Court did not elaborate. *Id.*

³⁰ See *id.* at 593 ("Submission [of a proposition] to the scrutiny of the scientific community is a component of 'good science,' in part because it increases the likelihood that substantive flaws in methodology will be detected.").

³¹ *Id.* at 594.

³² See, e.g., Michael C. Polentz, *Post-Daubert Confusion with Expert Testimony*, 36 SANTA CLARA L. REV. 1187, 1213 (1996) (noting that "the general acceptance test has the propensity to dominate the threshold gatekeeper inquiry"). Although the Court acknowledged that publication "is not a *sine qua non* of admissibility," it gave the factor unwarranted significance. *Id.* at 593. As one scientist explained, "The mere fact that research reports are published, even in the most prestigious journals, is no guarantee of their quality." PETER G. GOLDSCHMIDT & THEODORE COLTON, THE QUALITY OF MEDICAL LITERATURE: AN ANALYSIS OF VALIDATION ASSESSMENTS IN MEDICAL USES OF STATISTICS 370-91 (J.C. Bailor & F. Mosteller eds., 1986). By discussing publication and peer review—which are forms of consensus—as something other than and separate from consensus, the Court risked subsuming its other inquiries into the single issue of consensus. Indeed, despite the dispatch of *Frye*, courts continue to place a disproportionate emphasis on consensus. See, e.g., *United States v. Sherwood*, 98 F.3d 402, 408 (9th Cir. 1996) (upholding admission of fingerprint evidence because technique was generally accepted, peer reviewed and published, and indicating that not all *Daubert* factors must be considered in every case); *Hall v. Baxter Healthcare Corp.*, 947 F. Supp. 1387, 1406 (D. Or. 1996) (finding testimony inadmissible because unreviewed

peer review, publication, and general acceptance in the scientific community clearly affect scientists' evaluation of each others' work, these factors may have little, if anything, to do with the intrinsic merits of the research.³³ If a scientist's work looks otherwise sound, there is no a priori reason to dismiss it just because it has not found favor among the scientist's peers. Similarly, publication, peer review, and general acceptance do not validate research whose logical underpinnings are shaky. Many, if not most, articles published, after having received peer review and a measure of consensus, are later proven to be wrong.³⁴

4. *Error Rate.* The known or potential error rate of a particular scientific technique, and "the existence and maintenance of standards controlling the technique's operation"³⁵ are crucial inquiries.³⁶ Without knowing how often a given procedure yields mistaken results, it is difficult to assess either the reliability of the procedure or the validity of the inferences drawn from it.³⁷ Unless there are standards maintained and observed for the operation of a

based on notion that "[p]eer review and publication weigh heavily in the calculus of the reliability of expert testimony because such peer review 'increases the likelihood that substantive flaws in methodology will be detected' " (citing *Daubert*, 509 U.S. at 593)).

³³ Publication in science is competitive. Work may remain unpublished and achieve little acceptance in the scientific community, even though it has scientific merit, simply because it does not reflect contemporary interests. One recent example is the research of Dr. Kilmer S. McCully in the field of arteriosclerosis, which went scorned and unpublished not because of any intrinsic flaw in his research, but because he concluded that a vitamin deficiency rather than cholesterol alone was responsible for arteriosclerotic plaques. See Michelle Stacey, *The Fall and Rise of Kilmer McCully*, N.Y. TIMES, Aug. 10, 1997, at 25. This conclusion was unpopular among scientists, who wanted to alert the public to the dangers of high cholesterol, and to the pharmaceutical companies, who wanted to sell their anti-cholesterol drugs. Many years after his research was completed, and after a lifetime of being castigated by his colleagues, Dr. McCully is now beginning to achieve recognition for his pathbreaking research.

³⁴ See GOLDSCHMIDT & COLTON, *supra* note 32, at 370-91 (reporting that relatively few journal articles are scientifically sound in terms of reporting usable data and providing even moderately strong support for their inferences).

³⁵ *Daubert*, 509 U.S. at 594.

³⁶ See Shayne C. Gad & Carrol S. Weil, *Statistics for Toxicologists*, in PRINCIPLES AND METHODS OF TOXICOLOGY 221, 224 (A. Wallace Hayes ed., 3d ed. 1994) [hereinafter PRINCIPLES AND METHODS] (describing methodologies used in evaluating toxicological data).

³⁷ The absence of any error analysis is a characteristic and persistent flaw in much of what is proffered as criminal scientific evidence, undermining its validity. See Erica Beecher-Monas, *Blinded by Science: How Judges Avoid the Science in Scientific Evidence*, 71 TEMP. L. REV. 55 (1998); Paul C. Gianelli, "Junk Science": *The Criminal Cases*, 84 J. CRIM. L. & CRIMINOLOGY 105 (1993).

given technique, there can be no quality control, making a reliability assessment impossible.³⁸

It is therefore striking that the Court gave no guidance as to how a judge is to determine or evaluate the error rate or to evaluate controls or possible confounding factors.³⁹ Even where experts provide an error analysis, and good scientists normally provide an error analysis with their results, the type of analysis may vary widely.⁴⁰ Different scientific disciplines have different ways of estimating and controlling errors.⁴¹ The Court neither acknowledged those differences nor gave the lower courts any clues about how they should be assessed. Finally, although the Supreme Court correctly recognized that statistical concepts are important in assessing the validity of scientific conclusions, it did little more than mention the issue. The Court's cursory treatment of this factor made it far too easy for trial judges to ignore and gave too little guidance to those who wish to effectuate its purpose.

B. *DAUBERT* AND *JOINER* ON THE QUESTION OF "FIT"

The Court's express intention in *Daubert* was to switch the trial courts' focus from whether the expert's conclusions had garnered a scientific consensus to whether the expert's techniques and

³⁸ See KARIN D. KNORR-CETINA, *THE MANUFACTURE OF KNOWLEDGE: AN ESSAY ON THE CONSTRUCTIVIST AND CONTEXTUAL NATURE OF SCIENCE* 21 (1981) ("[T]he major task of the laboratory is to rule out possibilities, manipulate the balance of choices so that one becomes more attractive than the others, and to up- or down-grade variables with respect to alternatives."); FRANK C. LU, *BASIC TOXICOLOGY: FUNDAMENTALS, TARGET ORGANS, AND RISK ASSESSMENT* 73-86 (1996) (discussing acquisition of data in acute, short-term, and long-term toxicity studies).

³⁹ Controls are unexposed to the agent or procedure in question; confounding factors are variables related to both the exposure and the result which can therefore obscure the relationship between the agent and the result. See Bernard D. Goldstein & Mary Sue Henifin, *Reference Guide on Toxicology*, in *FEDERAL JUDICIAL CTR.*, *supra* note 8, at 181, 214.

⁴⁰ See KENNETH R. FOSTER & PETER W. HUBER, *JUDGING SCIENCE: SCIENTIFIC KNOWLEDGE AND THE FEDERAL COURTS* 70 (1997) (discussing various ways to classify scientific error). As noted above, an exception to this is found in the criminal context, where experts seldom proffer any error analysis with their studies. See *supra* note 37.

⁴¹ For a concise explanation of how statistics are used in scientific evidence and a description of statistics as "the science of uncertainty, a body of methods for making inferences and decisions when faced with fallible observations or other forms of uncertainty," see Stephen E. Fienberg et al., *Understanding and Evaluating Statistical Evidence in Litigation*, 36 *JURIMETRICS J.* 1, 3 (1995).

methodology were valid.⁴² That does not mean, as the Court later explained in *Joiner*, that the trial judge could ignore the expert's conclusions.⁴³ Rather, the trial court must examine the expert's methodology and techniques for consistency with the expert's conclusions and with the facts of the case at hand.⁴⁴ Conclusions and methodology, the Court pointed out in *Joiner*, are not entirely distinct from one another, and there must be a valid connection between them.⁴⁵

The Supreme Court's concern with "fit" is well founded in scientific theory. Observation and interpretation, and the connection between them, are concerns basic to scientific reasoning.⁴⁶ Even if the theory is testable, and the testing was performed correctly with the proper controls, the question is whether the experiments really show what they are intended to show.⁴⁷ In neither *Daubert* nor in *Joiner*, however, did the Court give much guidance on how to resolve these inherent problems.

The sole example of "fit" in *Daubert* was the relevance of a study of the phases of the moon, which the Court explained would be relevant to whether a particular night was dark if darkness was a fact in issue but not to whether an individual was likely to have behaved irrationally on that night.⁴⁸ The Court's explanation of this

⁴² Initially, the Supreme Court in *Daubert* recognized that the relevance of expert testimony depends upon whether the proffered testimony will resolve issues presented in the legal dispute before the court. *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 591 (1993). The role of the court, noted the majority, is to resolve factual disputes, some of which may involve contested scientific evidence. *Id.* The key question here is whether the proposed testimony is based on data which "fits" or is validly connected to the facts of the case. *Id.*

⁴³ *General Elec. Co. v. Joiner*, 522 U.S. 136, 146 (1997).

⁴⁴ *Id.*

⁴⁵ *Id.*

⁴⁶ Epidemiological studies, for example, lack the laboratory control over factors which may affect the quality of data. Exact determination of exposure levels, for example, is generally impossible in epidemiological studies, and control over other environmental factors may be difficult. In toxicology studies, the extrapolation of results from animal studies to humans may be problematic depending on, among other things, the type of animal used. KNORR-CETINA, *supra* note 38, at 21.

⁴⁷ For example, epidemiologists have developed a set of standards to take these variables into account. See generally DAVID G. KLEINBAUM ET AL., *EPIDEMIOLOGIC RESEARCH: PRINCIPLES AND QUANTITATIVE METHODS* (1982).

⁴⁸ *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 591 (1993). What is valid in one context to resolve one legal dispute, the Supreme Court observed, may not be valid in another. *Id.*

key concept in the *Joiner* opinion was less cursory but more cryptic. In upholding the trial court's exclusion of the evidence in *Joiner* and reiterating that the abuse of discretion standard applies to *Daubert* determinations as well as to other evidentiary rulings, the Supreme Court reviewed each of the four proffered epidemiology studies and two animal studies individually and found that none of them "fit" the plaintiff's experts' conclusions that Joiner's exposure to polychlorinated biphenyls (PCBs) was an aggravating factor in his developing lung cancer.⁴⁹ The opinion fails to explain, however, just what would make the kind of animal studies that scientists normally perform relevant to a toxic tort case. It also failed to explain how an expert may properly rely on numerous studies that build on each other without having the studies sequentially excluded as separately inadequate and therefore failing the "fit" of the case.

The two animal studies in *Joiner* involved infant mice that were injected with PCBs through the abdomen.⁵⁰ These studies not only varied factually,⁵¹ but the plaintiff's experts offered neither any studies to show any link between cancer and PCBs exposure in *adult* mice nor any evidence that any other species exposed to PCBs developed cancer.⁵² Thus, according to the Court, the problem was not the validity of the studies per se but the failure to provide evidence to link the factually distinguishable studies to the plaintiff's disease.⁵³ Because the experts failed to perfect the analogy

⁴⁹ *Joiner*, 522 U.S. at 143-45. The foundation for the experts' proffered testimony consisted of two animal studies and four epidemiological studies. Chief Justice Rehnquist (who had dissented in *Daubert*), writing for the majority, explained that the trial court did not abuse its discretion either in finding that "the animal studies on which respondent's experts relied did not support his contention that exposure to PCBs had contributed to his cancer" or in finding that the epidemiology studies on which the plaintiff's experts relied did not "fit" the conclusions which the experts espoused. *Id.* at 144.

⁵⁰ *Id.*

⁵¹ The factual differences between Joiner's exposure and the experiments involved the studies' use of infant mice, whereas Joiner was an adult human; the mice were subjected to massive doses of highly concentrated PCBs injected directly into their abdomens, whereas the plaintiff was exposed to much smaller PCB concentrations through his skin and lungs. Also, the plaintiff developed different types of tumors from the mice. *Id.*

⁵² *Id.*

⁵³ *Id.*

they sought to make, their testimony did not meet the requirements of "fit."⁵⁴

The Court found that the four epidemiology⁵⁵ studies were equally flawed.⁵⁶ Not one of the studies concluded definitively that PCBs exposure increased the incidence of cancer in the exposed population.⁵⁷ Two studies involving workers exposed to PCBs noted a higher mortality rate among these workers than expected, but neither could demonstrate a link between the increased mortality rate and exposure to PCBs.⁵⁸ A third study of workers exposed to mineral oil showed a statistically significant increase in lung cancer deaths in these workers, but the mineral oil was not shown to contain PCBs and the study expressly limited its conclusions to the type of mineral oil studied.⁵⁹ The fourth study involved Japanese subjects who had been exposed to other carcinogens in addition to PCBs in the toxic rice oil which they had eaten.⁶⁰ Because none of the proffered studies individually supported a conclusion that PCBs increase the risk of cancer, the Court found they did not "fit" the underlying hypothesis that plaintiffs' cancer was aggravated by his exposure to PCBs.⁶¹

Thus, although the Court found nothing wrong with the principles or methodology underlying the individual studies, it found that each of them failed to support the expert's theory that exposure to PCBs aggravated the development of the plaintiff's cancer.⁶² The majority recognized that there might be an argument that the combined evidence provided support for the experts, but remarked that the plaintiff offered neither evidence to show how the studies were analytically linked nor guidance as to the cumulative impact of the studies.⁶³ The Court acknowledged that "[t]rained experts

⁵⁴ *Id.* at 144-45.

⁵⁵ Epidemiology is the "study of the occurrence and distribution of disease among people." Goldstein & Henifin, *supra* note 39, at 215.

⁵⁶ *Joiner*, 522 U.S. at 145-46.

⁵⁷ *Id.* at 145.

⁵⁸ *Id.*

⁵⁹ *Id.* at 145-46.

⁶⁰ *Id.* at 146.

⁶¹ *Id.* at 145.

⁶² *Id.*

⁶³ *Id.* at 144-45.

commonly extrapolate from existing data," recognizing that experts may need to rely on a number of imperfect studies to reach a conclusion.⁶⁴ The problem with the *Joiner* plaintiff's expert testimony, in the Court's view, was that the experts failed to explain their extrapolations.⁶⁵ With respect to the animal studies, the Court explained that the issue was not whether animal studies could ever show causation in humans but "whether these experts' opinions were sufficiently supported by the animal studies on which they purported to rely."⁶⁶

The *Joiner* Court took an overly-mechanistic view of causation in upholding the district court's finding that the cancer of a thirty-five-year-old former smoker with a family history of early deaths from lung cancer had not been causally linked by the proffered evidence of PCBs exposure.⁶⁷ After all, everyone knows that smoking causes lung cancer, and everyone knows that genetics are also important determinants of lung disease. *Joiner* had both of these risk factors working against him. Using common sense and the leeways of the legal tradition, it is not hard to see why the district court decided *Joiner* had no case despite undisputed exposure to toxins.⁶⁸ He was doomed anyway.

The district court's conclusion may reflect common sense, but it demonstrates little understanding of probabilistic evidence.⁶⁹ The Supreme Court also appears to require precisely the kind of mechanistic causation that science has long since abandoned.⁷⁰

⁶⁴ *Id.* at 146.

⁶⁵ *Id.* at 146-47.

⁶⁶ *Id.* at 144.

⁶⁷ *Id.* at 144-46.

⁶⁸ *Joiner v. General Elec. Co.*, 864 F. Supp. 1310, 1324 (N.D. Ga. 1994), *rev'd* 78 F.3d 524 (11th Cir. 1996), *rev'd* 522 U.S. 136 (1997).

⁶⁹ See *infra* notes 276-317 and accompanying text (discussing causation theories).

⁷⁰ See, e.g., Troyen A. Brennan & Robert F. Carter, *Legal and Scientific Probability of Causation of Cancer and Other Environmental Disease in Individuals*, 10 J. HEALTH POL., POL'Y & L. 33, 34 (1985) (discussing inadequacies of common law in dealing with causation of injuries related to exposure to hazardous substances).

Probabilistic evidence is inherently uncertain,⁷¹ yet scientists routinely accept these uncertainties.⁷²

How the Court rationalized upholding the result is even more troubling. Unable to say that Joiner was a dead man walking for fear of opening up tort doctrines such as last clear chance,⁷³ the Court engaged in sophistry about the relevance of the proffered evidence. The question arose on a motion for summary judgment, and the immediate question before the Court was whether the proffered evidence was scientifically valid. PCBs are indisputably linked to lung cancer.⁷⁴ Thousands of studies show this, forming the basis of the regulatory ban on PCBs.⁷⁵ Joiner was unquestionably exposed to PCBs.⁷⁶ The Court had a problem if it wanted to get rid of this case so it did so under the aegis of "fit" by refusing to acknowledge the interrelationship of the vast literature on PCBs (a great deal of which had been placed in evidence) with the individual studies proffered.⁷⁷ It did this by excoriating the experts and, by implication, plaintiffs' lawyers, for failing to explain how to link the literature with the studies, and how to extrapolate these studies from animals to humans and from the humans studied to Joiner.⁷⁸

⁷¹ *Id.* at 46 ("when a scientist states that a chemical causes cancer, the term *cause* is consciously hedged, and used in its most scientific manner").

⁷² For example, toxicologists routinely accept animal data and, in the absence of better data, *in vitro* studies of genetic mutations. *See, e.g.*, Philip E. Enterline, *Attributability in the Face of Uncertainty*, 78 CHEST 377, 379 (1980).

⁷³ Last clear chance is a tort doctrine originating in the mid-1800's and is illustrated by the English case in which the plaintiff left his ass tied up on the highway and the defendant drove into it. W. PAGE KEETON ET AL., PROSSER AND KEETON ON THE LAW OF TORTS § 66, at 463 (5th ed. 1984) (citing *Davies v. Mann*, 152 Eng. Rep. 588 (1842)). Multiple causation is an issue that has troubled the courts since well before the upsurge in toxic tort litigation. *Id.* §§ 41-42, 65-66. The textbook rule is that liability may be imposed only if a tortious act was a substantial factor in producing injury. *Id.* § 41, at 271. Determining whether a given causal link constitutes a substantial factor is itself a probabilistic inquiry that has occasioned much controversy in the courts. *Id.* § 66, at 462-63.

⁷⁴ Joiner's experts relied on the "studies of at least thirteen different researchers, and referred to several reports of the World Health Organization," all of which implicated PCBs in cancer formation. *General Elec. Co. v. Joiner*, 522 U.S. 136, 151 (1997) (Stevens, J., concurring in part and dissenting in part).

⁷⁵ *Id.* at 139.

⁷⁶ *Id.*

⁷⁷ *Id.* at 144-46.

⁷⁸ *Id.* at 144.

Along the way, the Court said all the right things. It acknowledged the admissibility of animal studies for human causation, and then it found *these* animal studies deficient, explaining that Joiner's experts failed to explain the extrapolations properly.⁷⁹ The Court acknowledged that studies ought to be interpreted in conjunction with each other, but then it found that Joiner's experts failed to explain how *these* studies could be linked.⁸⁰ According to the Court, the record was simply deficient—it was a lawyering failure.⁸¹ As the dissent points out, however, a perusal of the documents before the district court reveals that a great deal more information was submitted in the *Daubert* hearing than ever made it into either the judge's opinion or the parties' record.⁸²

Disposing of the case under the aegis of admissibility permitted the Court to avoid the more troubling issue of whether someone who has put himself in harm's way can claim that his death resulted from another equally fatal cause. Can a man playing Russian roulette be murdered when someone else shoots him at the same time that his own gun went off? It is a tough question, and one that a jury might have trouble with, but that does not excuse the Court's intellectual dishonesty. Had it reached a jury, the proffered evidence might have been difficult to weigh. That should not, however, excuse the Court's circumvention of the jury under the pretext of admissibility.

At its most scientifically defensible, the Court's opinion can be read to articulate two basic principles. First, experts may rely on the cumulative evidence of numerous studies, none of which is precisely analogous, to form the basis of their opinions.⁸³ The Court found that a single dispositive study is not required, nor is each study relied on required to support the whole of the expert's conclusion.⁸⁴ Despite this pronouncement, the Court excluded the individual studies piecemeal rather than examining the data in their entirety.

⁷⁹ *Id.* at 144-45.

⁸⁰ *Id.*

⁸¹ *Id.*

⁸² *Id.* at 151 (Stevens, J., concurring in part and dissenting in part).

⁸³ *Id.* at 146-47.

⁸⁴ *Id.*

Second, the Court found that animal studies were not categorically excludable.⁸⁵ The Court acknowledged the importance of animal studies to toxicology.⁸⁶ It argued that these studies could be considered in conjunction with other available data as long as the expert's extrapolations were explicit and valid.⁸⁷ Nonetheless, the Court excluded the animal studies without offering any guidance as to how a trial judge is to determine when and how experts may extrapolate and the limitations that ought to be imposed on such extrapolations.⁸⁸ Indeed, its discussion was limited to remarking on the analytical gaps between the proffered toxicology and epidemiology studies and the experts' opinions, with no explanation of how the gaps should have been filled.⁸⁹ As a result, the issue of "fit" will undoubtedly continue to plague the courts, particularly with respect to cumulative scientific evidence.

C. THE TROUBLED PROGENY OF *DAUBERT* AND *JOINER*

Although the Supreme Court opinions in *Daubert* and *Joiner* introduced judges to some basic scientific concepts, they left a great many questions unanswered. How a judge determines whether a particular experimental method is good science or how a judge decides whether one type of study is superior to another are open questions.⁹⁰ The Supreme Court also failed to acknowledge that the scientific method is not the same for all scientific disciplines and that scientific reliability is not an all-or-nothing proposition, but rather depends on the application of the evidence and the acceptable risk of error.⁹¹ Neither *Daubert* nor *Joiner* specifically addressed the way in which separate studies, each individually inadequate to support a given conclusion, may cumulatively provide foundation for

⁸⁵ *Id.* at 144-45.

⁸⁶ *Id.*

⁸⁷ *Id.*

⁸⁸ *Id.*

⁸⁹ *Id.*

⁹⁰ For a critique of the indeterminacy of the Supreme Court's opinion in *Daubert*, see generally Randolph N. Jonakait, *The Meaning of Daubert and What That Means for Forensic Science*, 15 CARDOZO L. REV. 2103 (1994).

⁹¹ See Paul S. Milich, *Controversial Science in the Courtroom: Daubert and the Law's Hubris*, 43 EMORY L.J. 913, 920 (1994) (urging caution in admitting novel scientific evidence).

a scientifically valid expert opinion. Statistical analyses, although concededly important to a *Daubert* determination, were neither discussed nor explained. Probabilistic causation issues were ignored.

As a result, post-*Daubert* examples of courts' mishandling of scientific evidence abound.⁹² This mishandling of scientific evidence is particularly acute in toxic tort cases where proving causation nearly always involves the use of scientific experts. The courts' analyses frequently founder on several major issues. These include the admissibility and evaluation of animal studies, the impact of cumulative studies, and the use of statistical analysis.

1. *Animal Studies.* In broad terms, animal studies are performed by exposing animals to a particular chemical and extrapolating the results to humans using what is known about the structure, function, and metabolism of the particular chemical and the similarity of the actions of the chemical in the animals studied and human beings.⁹³ Even under the old *Frye* general acceptance test,⁹⁴ and despite near unanimity in the scientific community that animal studies are reliable indicators and useful predictors of human disease,⁹⁵ courts have had enormous difficulty accepting the validity

⁹² See generally Paul S. Miller & Bert W. Rein, *Whither Daubert? Reliable Resolution of Scientifically-Based Causality Issues in Toxic Tort Cases*, 50 RUTGERS L. REV. 563, 563 (1998) (arguing that courts should decide causation rather than juries). In criminal cases, the issues that plague the courts concern the absence of empirical bases for a wide range of purportedly scientific identification techniques. Margaret A. Berger, *Procedural Paradigms for Applying the Daubert Test*, 78 MINN. L. REV. 1345, 1354-55 (1994). In civil cases other than toxic torts, involving expert testimony from engineers and accountants for example, the question often surfaces as to whether such evidence should be scrutinized for validity under *Daubert* guidelines. Recently, in *Kumho Tire Co. v. Carmichael*, 67 U.S.L.W. 4179 (U.S. Mar. 23, 1999), the Supreme Court held that expert, non-scientific engineering testimony is also subject to *Daubert* analysis.

⁹³ See Proposed Guidelines, *supra* note 11, at 17,974-77; see also D. Cooper Rees & Dale Hattis, *Developing Quantitative Strategies for Animal to Human Extrapolation*, in PRINCIPLES AND METHODS, *supra* note 36, 275-315 (presenting at guidelines for extrapolating animal toxicity data to humans).

⁹⁴ See *Frye v. United States*, 293 F. 1013, 1014 (D.C. Cir. 1923) (holding that only "generally accepted" scientific evidence is admissible).

⁹⁵ See, e.g., Arnold L. Brown, *The Meaning of Risk Assessment*, 37 J. ONCOLOGY 302 (1980) (expressing doubt that anyone "can seriously question the relevance of animal data to the human experience"); I. Bernard Wasserstein, *Mitogenesis is Only One Factor in Carcinogenesis*, 251 SCIENCE 387, 388 (1991) (describing rodent bioassays as an invaluable tool in assessing carcinogenesis in humans.).

of testimony that extrapolates evidence about human injuries from animal studies.⁹⁶ Post-*Daubert* courts often find expert testimony based primarily on animal studies inadmissible simply because they relate to animals rather than to humans.⁹⁷ Nonetheless, animal studies are routinely relied on by scientists in determining toxicity.⁹⁸ Their wholesale rejection by the courts is unwarranted and reflects the courts' ignorance of basic scientific precepts.

Moreover, although epidemiologic studies may be the preferred proof of general causation, they are no panacea.⁹⁹ Indeed, animal studies have a number of advantages: laboratory conditions permit

⁹⁶ See, e.g., *In re Selwood v. Oxford Chems. Inc.*, No. 91-5619 (3d Cir. Feb. 21, 1992) (finding inadmissible expert testimony regarding health effects of hazardous substances based only on animal studies); *Bell v. Swift Adhesives Inc.*, 804 F. Supp. 1577, 1581 (S.D. Ga. 1992) (finding expert testimony based on animal studies insufficient to withstand motion for summary judgment); *Ball v. Joy Mfg. Co.*, 755 F. Supp. 1344, 1348 (S.D. W. Va. 1990) (finding expert testimony on toxic effects of PCBs inadmissible and noting that "jurisdictions are in disagreement over the admissibility of expert medical or scientific testimony based on animal studies where the reliability of such evidence is contested"); *In re Agent Orange Prod. Liab. Litig.*, 611 F. Supp. 1223, 1241 (E.D.N.Y. 1985) (finding animal studies "of so little probative force and . . . so potentially misleading as to be inadmissible"). Another example is the failure of the court in *Stites v. Sundstrand Heat Transfer, Inc.*, 660 F. Supp. 1516, 1521 (W.D. Mich. 1987), to see the relevance of proffered data on metabolic transformation and carcinogenesis of trichloroethane ("TCE"). Although the plaintiff, who sought damages for exposure to TCE on a risk of cancer theory, offered expert testimony on the metabolism of TCE, the mechanism by which exposure to TCE "can increase an individual's susceptibility to cancer," and on TCE's similarity to the molecular structure of known human carcinogens, the court granted summary judgment to the defendant. *Id.* at 1521, 1526. This result seems to embody an unnecessarily restrictive view of the certainty level demanded before evidence will be deemed sufficient. See, e.g., *Raynor v. Merrell Pharms., Inc.*, 104 F.3d 1371, 1375 (D.C. Cir. 1997) (requiring human experiments or epidemiological data); *Wade-Greaux v. Whitehall Labs., Inc.*, 874 F. Supp. 1441, 1485 (D.V.I. 1994) (finding human studies necessary for admissibility), *aff'd*, 46 F.3d 1120 (3d Cir. 1994). In spite of its dispositive-sounding dicta about animal studies, the court in *Wade-Greaux* actually examined the proffered studies and soundly critiqued them. *Wade-Greaux*, 874 F. Supp. at 1483-84; see also Margaret A. Berger, *Eliminating General Causation: Notes Towards a New Theory of Justice and Toxic Torts*, 97 COLUM. L. REV. 2117, 2124-25, 2145 (1997) (noting courts' difficulties with admitting expert testimony based on animal studies and proposing shifting burden of proof to defendants in toxic torts cases).

⁹⁷ See, e.g., *Raynor*, 104 F.3d at 1375 (requiring human experiments or epidemiological data); *Hall v. Baxter Healthcare Corp.*, 947 F. Supp. 1387, 1397 (D. Or. 1996) (finding animal studies inadmissible to prove causation in humans unless there are "good grounds to extrapolate from animals to humans"); *Wade-Greaux*, 874 F. Supp. at 1430 (requiring human studies for admissibility); *Merrell Dow Pharms., Inc. v. Havner*, 953 S.W.2d 706, 729-30 (Tex. 1997) (rejecting animal studies as unreliable under Texas equivalent of *Daubert*).

⁹⁸ Berger, *supra* note 96, at 2124-25, 2145.

⁹⁹ *Id.* at 2125.

the researcher to have better control over the experimental conditions; high dosages more accurately reveal the presence of a dose-response relationship; and the time required to conduct an animal study is much shorter.¹⁰⁰ Where complex exposures are involved in epidemiological investigations—as in most toxic torts—it is difficult to define the causal element.¹⁰¹ Moreover, designing and conducting a sound epidemiologic study is difficult, time-consuming, and expensive.¹⁰² As a result, such studies are in short supply.¹⁰³ Because there are so few good epidemiologic studies available, animal studies are often the primary source of information regarding the health effects of chemicals.¹⁰⁴

Despite the obvious differences between humans and laboratory animals, biochemical and metabolic processes carried out in most organs are similar.¹⁰⁵ All of the forty or so chemicals recognized to cause cancer in humans also cause cancer in animals.¹⁰⁶ There is

¹⁰⁰ See Brennan & Carter, *supra* note 70, at 44 (stating that epidemiologic studies are more dependent on “probabilistic-type reasoning” than animal studies); 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, 654 (1992) (noting advantages of animal studies over epidemiologic studies).

¹⁰¹ See Barbara D. Beck et al., *The Use of Toxicology in the Regulatory Process*, in PRINCIPLES AND METHODS, *supra* note 36, at 19, 23-24.

¹⁰² Berger, *supra* note 96, at 2127-28. Controls among human populations are a more significant problem than they are under laboratory conditions. See Troyen A. Brennan, *Causal Chains and Statistical Links: The Role of Scientific Uncertainty in Hazardous-Substance Litigation*, 73 CORNELL L. REV. 469, 507 (1988) (noting that researchers cannot control conditions of epidemiologic studies as well as they can in laboratory experiments).

¹⁰³ See Landau & O’Riordan, *supra* note 12, at 532, 565-66 (concluding that despite scarcity of epidemiologic studies, animal studies should not be admissible, a conclusion with which this Article takes issue).

¹⁰⁴ See *id.* at 534. Indeed, for most toxic tort cases, there is a paucity of data, not necessarily because there is no causation but because many toxic tort cases share a history of manufacturers’ neglect in conducting basic safety research. See Rebecca S. Dresser et al., *Breast Implants Revisited: Beyond Science on Trial*, 1997 WIS. L. REV. 705, 732-34 (detailing “repetitive pattern of manufacturer neglect” of basic safety research in asbestos, tobacco, Agent Orange, Dalkon Shield, ultra-absorbent tampons, and Bendectin litigation).

¹⁰⁵ See Beck et al., *supra* note 101, at 38; Carl F. Cranor et al., *Judicial Boundary Drawing and the Need for Context-Sensitive Science in Toxic Torts After Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 16 VA. ENVTL. L.J. 1, 51 nn.224-27 (1996) (quoting D.P. Rall et al., *Alternatives to Using Human Experience in Assessing Health Risks*, 8 ANN. REV. PUB. HEALTH 335, 356 (1987) for the proposition that “there are more physiologic, biochemical, and metabolic similarities between laboratory animals and humans than there are differences”).

¹⁰⁶ International Agency for Research on Cancer, *Preamble*, in 63 IARC MONOGRAPHSON THE EVALUATION OF CARCINOGENIC RISKS TO HUMANS 9, 17 (1995).

typically a close correspondence between the target organ in these studies and at least one of the animal species studied.¹⁰⁷ That is not to say that differences in metabolism, body size, dose, lifespan, and other factors should not be considered. Complex issues do exist, relating to route of exposure, dose-response models, scaling factors, and cross-species variations.¹⁰⁸ Nonetheless, with all their attendant uncertainties, scientists routinely rely on these studies, understanding them as reliable probabilistic statements about causality.¹⁰⁹ Courts should not, therefore, categorically reject animal studies.¹¹⁰

In addition to the wholesale exclusion of animal studies, courts also appear to founder on extrapolating the results from animal studies to humans.¹¹¹ The utility of animal studies designed for

¹⁰⁷ *Id.*; see also James Huff, *Chemicals and Cancer in Humans: First Evidence in Experimental Animals*, 100 ENV'TL HEALTH PERSP. 201, 204 (1993) ("the array and multiplicity of carcinogenic processes are virtually common among mammals").

¹⁰⁸ See Brennan & Carter, *supra* note 70, at 42-44 (discussing complexity of animal bioassay testing).

¹⁰⁹ See *id.* ("Scientists assume that substances which are carcinogenic in animals are carcinogenic in humans—an assumption which has been proven time and again when doing animal bioassays of known human carcinogens.").

¹¹⁰ Despite the major uncertainties in extrapolating (or "scaling") an equivalent dose in humans, most toxicologists believe that animal studies are reliable predictors of human toxicity. See NATIONAL RESEARCH COUNCIL, RISK ASSESSMENT IN THE FEDERAL GOVERNMENT: MANAGING THE PROCESS 22 (1983) (premise that results from animal experiments are applicable to humans "underlies much of perimeutal biology and medicine").

¹¹¹ For example, in *Nelson v. American Sterilizer Co.*, 566 N.W.2d 671, 676 (Mich. Ct. App. 1997), the court excluded expert testimony that plaintiff's liver disease was caused by exposure to chemicals used in sterilizing medical equipment merely because the animal studies the expert relied on were high dose rather than low dose studies. This appears to be a fundamental misconception about the usefulness of animal studies. However, this may be another lawyering failure, as the court does point out that plaintiff's experts provided no understandable scientific basis for their extrapolations. The failure of the plaintiff in *Joiner* to provide the basis for extrapolating the results of the animal tests also appears to be the basis of the *Nelson* court's exclusion. The court responded to the plaintiff's argument that it should not categorically exclude animal studies by pointing out the failure of the expert to draw a convincing analogy from these studies to human causation. *Id.*; see General Elec. Co. v. *Joiner*, 522 U.S. 136, 144-45 (1997) (rejecting plaintiff's expert testimony because opinion was seemingly far removed from animal studies). In the *Paoli* PCB litigation, the Third Circuit reinstated the excluded animal studies despite defendants' argument that "test animals are often very sensitive to chemicals due to . . . physiological, biological and metabolic pathways which are different than those of humans." *In re Paoli RR Yard PCB Litig.*, 35 F.3d 717, 779 (3d Cir. 1994). In *Paoli*, unlike in *Joiner*, the expert was able to testify that the animals studied (monkeys, in *Paoli*) had sensitivities to PCBs similar to humans. *Id.* at 779-80.

regulatory risk assessment in toxic tort litigation is the subject of ongoing legal debate.¹¹² Physiological differences between the animals studied and humans (or as one court remarked, "humans are not rats"¹¹³), the high dosages ordinarily used in such studies in contrast to the low dosages typically at issue in toxic tort cases, and laboratory conditions are all contentious areas of litigation. Although there is some validity to these objections, they are not dispositive.¹¹⁴ High-dosage extrapolations from animals to humans are commonly used by toxicologists to provide realistic indications of causal relationships in humans. This is a cornerstone of scientific research.¹¹⁵ While there are always a few scientists who urge caution in the wholesale adoption of animal studies, this is a distinct minority. No modern scientist seriously questions their relevance, as long as the animal studies are supplemented by information regarding chemical structure, the manner of chemical absorption, metabolization, and distribution in the body, and as long as there is

¹¹² See Erin K.L. Mahaney, *Assessing the Fitness of Novel Scientific Evidence in the Post-Daubert Era: Pesticide Exposure Cases as a Paradigm for Determining Admissibility*, 26 ENVTL. L. J. 1161, 1183 (1996) (noting that debate centers around "variability in laboratory conditions, the variety of responses observed within the same and different species of animals, the administration of extremely high doses, the inherently arbitrary selection of low-dose extrapolation models, and the unreliability of interspecies extrapolation"). Courts sometimes categorically reject agency findings for purposes of tort litigation as having been developed for regulatory (preventive) purposes rather than litigation and thus needing a lower standard of proof. See, e.g., *Allen v. Pennsylvania Eng'g Corp.*, 102 F.3d 194 (5th Cir. 1996) (rejecting EPA's "weight of the evidence" methodology). There is little rationale for this bias against hazard characterization, as opposed to level of exposure methodology, which may have a preventive bias, since EPA carcinogenic risk assessment explicitly characterizes the level of hazard gleaned from all the biological information available as "known/likely," "cannot be determined," or "not likely" to be carcinogenic to humans. See Proposed Guidelines, *supra* note 11, at 17,961. Moreover, the EPA uses the same "more probable than not" standard to reach its risk assessments as is used in civil litigation. See *id.* I am not arguing that EPA risk assessments should be adopted as dispositive in toxic tort litigation. Rather, I suggest that there is nothing inherently biased about the methodology used in reaching the risk characterization. Cf. Ellen Relkin, *The Sword or the Shield: Use of Governmental Regulations, Exposure Standards and Toxicological Data in Toxic Tort Litigation*, 6 DICK. J. ENVTL. L. & POL'Y 1, 4 (1997) (cautioning against wholesale adoption of risk assessments models).

¹¹³ *International Union v. Pendergrass*, 878 F.2d 389, 394 (D.C. Cir. 1989).

¹¹⁴ See Beck et al., *supra* note 101, at 38 (discussing cross-species extrapolation).

¹¹⁵ *Id.* at 22-23.

a physiologic analogy in the species studied.¹¹⁶ The courts, however, have not yet caught up with this understanding and often mechanically reject testimony based on animal studies.

2. *Cumulative Impact of Scientific Evidence.* Science is a collaborative enterprise, and scientific studies simply cannot be interpreted in isolation. Scientific studies build upon one another. Because they ignore this pivotal idea, courts frequently find the issue of "fit" a difficult determination to make.¹¹⁷ Although the relevance of a single study may be a fairly straightforward determination, relevance becomes more complicated when a number of studies are involved, each of which is only marginally relevant, but which together purport to form the basis of an expert's conclusions. This is by far the most common situation, particularly in toxic tort cases. Only rarely will an expert rely on a single study. More often, many different studies may—and, indeed, should—be offered cumulatively to support the expert's position, although no single study could do so on its own.

In *Joiner*, for example, the Eleventh Circuit reversed the district court's exclusion of the plaintiff's expert testimony on this basis.¹¹⁸ It explained that "[o]pinions of any kind are derived from individual pieces of evidence, each of which by itself might not be conclusive, but when viewed in their entirety are the building blocks of a perfectly reasonable conclusion. . . ."¹¹⁹ The Supreme Court did not address this view of the evidence, except to note that the district

¹¹⁶ See Proposed Guidelines, *supra* note 11, at 17,968; JAMES L. SCHARDEIN, CHEMICALLY INDUCED BIRTH DEFECTS 27 (1993) (observing that "not a single chemical exists that is teratogenic in humans that has not produced malformations in rodents" and noting that confirming studies in multiple species increases confidence in the results); International Agency for Research on Cancer, *supra* note 106, at 17 (stating that all chemicals recognized to cause cancer in humans also cause cancer in animals); see also *Paoli RR Yard PCB Litig.*, 35 F. 3d 717, 743 (3d Cir. 1997) (explaining the "fit" requirement that "in order for animal studies to be admissible to prove causation in humans, there must be good grounds to extrapolate from animals to humans, just as the methodology of the studies must constitute good grounds to reach conclusions about the animals themselves").

¹¹⁷ See *Habecker v. Clark Equip. Co.*, 36 F.3d 278, 290 (3d Cir. 1994) (excluding accident reconstruction evidence as irrelevant); *In re Hanford Nuclear Reservation Litig.*, No. CY-91-3015-AAM, 1998 WL 775340, at *139-40 (E.D. Wash. Aug. 21, 1998) (finding too great an analytical gap between proffered animal studies and causation issue for human plaintiffs).

¹¹⁸ *Joiner v. General Elec. Co.*, 78 F.3d 524, 532 (11th Cir. 1996), *rev'g* 864 F. Supp. 1310 (N.D. Ga. 1994), *rev'd*, 522 U.S. 136 (1997).

¹¹⁹ *Id.*

court did not abuse its discretion in finding that the individual studies proffered as the bases of the expert testimony could not support the experts' conclusions¹²⁰ and that Joiner did not explain how to link the studies together to show their applicability to his claims.¹²¹

3. *Statistical Analysis.* Probabilistic attribution and statistical analysis frequently confound the courts.¹²² Yet, statistical analysis and probabilistic thinking are key to understanding the validity of scientific studies. Indeed, probabilistic reasoning is the foundation of modern science.¹²³ Epidemiology is meaningless without statistics.¹²⁴ Statistical methods are crucial in testing hypotheses, constructing and using models, and reducing the number of variables in a system (by use of mean or standard deviation, for example).¹²⁵ The courts habitually misunderstand the difference between biological significance and statistical significance and the nature of different types of data.¹²⁶ If, for example, a few animals

¹²⁰ See *General Elec. Co. v. Joiner*, 522 U.S. 136, 144-45 (1997).

¹²¹ Joiner's brief to the Supreme Court merely argues that "[a]nalysis of the overall weight of the available data is the very methodology used by the EPA." Brief for Respondents at 40, *Joiner*, 522 U.S. 136 (No. 96-188), available in 1997 WL 436250. It may be, as Justice Stevens remarked in his concurrence and dissent, that there existed sufficient studies cumulatively to support the plaintiffs' position—the court of appeals pointed out that the plaintiffs' experts relied on 13 studies as well as reports of the World Health Organization—but only one of the studies made it into the record, and the district court discussed only the six studies referred to above. *Joiner*, 522 U.S. at 151 (Stevens, J., concurring in part and dissenting in part). The majority also remarked on Joiner's failure to explain how and why the experts could have extrapolated their opinions from the proffered studies. *Id.* at 144-45.

¹²² See, e.g., *In re Joint Asbestos Litig.*, 52 F.3d 1124, 1128-29 (2d Cir. 1995) (confusing statistical significance with magnitude of risk); see also Brennan & Carter, *supra* note 70, at 35 (observing that judges are "peculiarly hostile to the sort of probabilistic evidence of cancer causation that science can now deliver").

¹²³ See Brennan & Carter, *supra* note 70, at 38 (discussing influence of probabilistic reasoning in modern medical research).

¹²⁴ See GARY D. FRIEDMAN, *PRIMER OF EPIDEMIOLOGY* 1-3 (2d ed. 1980) (defining epidemiology).

¹²⁵ See Gad & Weil, *supra* note 36, 37, at 223 (discussing functions of statistics).

¹²⁶ See *id.* at 222 (explaining differences in types of significance); see, e.g., *Allen v. Pennsylvania Eng'g Corp.*, 102 F.3d 194, 198 (5th Cir. 1996) (rejecting testimony in absence of statistical significance); *Kelley v. American Heyer-Schulte Corp.*, 957 F. Supp. 873 (W.D. Tex. 1997) (rejecting testimony based on studies failing to show statistical significance); *LeBlanc v. Merrell Dow Pharms., Inc.*, 932 F. Supp. 782, 784 (E.D. La. 1996) (rejecting testimony due to lack of studies showing statistical significance); *Ambrosini v. Upjohn Co.*, No. 84-3483, 1995 WL 637650 (D.D.C. Oct. 18, 1995) (rejecting epidemiological testimony failing to demonstrate a statistically significant association), *rev'd* 101 F.3d 129 (D.C. Cir.

exposed to a particular substance develop a rare type of tumor, there may be biological significance without any statistical significance because of the paucity of data. Courts, ignorant of this factor, nearly always require data with a particular level of statistical significance.¹²⁷

Statistics are descriptive, and the parameters used for the description are subject to choice. Both the description of the location of the data and the measure of the dispersion about the location are chosen parameters. Courts, however, typically treat statistical statements as though they were immutable laws rather than descriptive statements.¹²⁸ Statistics, like other descriptive endeavors, are subjective, but courts treat statistical statements as though they were objective.

Another mistake that courts make is the categorical exclusion of epidemiological studies that do not demonstrate a relative risk of at least two.¹²⁹ The ostensible rationale for this exclusion is that a relative risk of two indicates that twice as many people in an exposed group contracted a disease as those who contracted the disease in an unexposed control group.¹³⁰ Courts, therefore, conclude that the affected people in the exposed group "more

1996); *Merrell Dow Pharms., Inc. v. Havner*, 953 S.W.2d 706, 724 (Tex. 1997) (rejecting under Texas equivalent of *Daubert* test, statistical significance levels other than five percent).

¹²⁷ See *General Elec. Co. v. Joiner*, 522 U.S. 136, 145-46 (1997) (affirming lower court's decision to exclude epidemiologic study for lack of statistical significance); *Hall v. Baxter Healthcare Corp.*, 947 F. Supp. 1387, 1405 (D. Or. 1996) (noting that studies did not support expert testimony because only one out of 16 found statistical relationship); *Havner*, 953 S.W.2d at 724 (requiring statistical significance level of five percent admissibility).

¹²⁸ Cf. Brennan & Carter, *supra* note 70, at 39 (explaining scientist's understanding that "causality is not a simple either/or proposition: the probability that one event caused another can be increased or decreased depending on how well new evidence fits with the guiding theory").

¹²⁹ See, e.g., *Daubert v. Merrell Dow Pharms., Inc.*, 43 F.3d 1311, 1320-21 (9th Cir. 1995) (excluding evidence of relationship between Bendectin and birth defects when studies found relative risks of less than two). Relative risk is the ratio of the risks for exposed and unexposed people; a relative risk of one means the same risk for exposed and unexposed people, while a relative risk of two indicates a doubling of the risk for exposed and unexposed people. Nancy A. Dreyer, *An Epidemiologic View of Causation: How it Differs From the Legal*, 61 DEF. COUNS. J. 40, 40 (1994).

¹³⁰ See Cranor et al., *supra* note 105, at 38 (arguing that it is "error to exclude epidemiological evidence simply because it reveals a relative risk less than two, unless there is no other supporting evidence").

probably than not" contracted their disease through exposure.¹³¹ In other words, courts excluding studies with relative risks less than two contend that unless the study shows it is twice as likely that the agent caused the disease, the study fails to meet the legal preponderance test.¹³² The requirement of a relative risk of two, however, rests on a misunderstanding of statistics,¹³³ and reflects a misinterpretation of the relative risk in terms of posterior probabilities.¹³⁴

In terms of posterior probabilities, while it may make sense to exclude evidence that cannot meet that standard if the plaintiffs are relying on a single study with a relative risk less than two, it makes little sense if there are several studies which should be considered in conjunction.¹³⁵ Atom bomb survivors, for example, exposed to ionizing radiation (a universally acknowledged carcinogen), show a relative risk of all cancers taken together (except leukemia) of less than two.¹³⁶ For courts prone to categorical exclusions, evidence of a type of cancer other than leukemia would be inadmissible. Such a result is absurd from a scientific viewpoint.

Meta-analysis of data (or reanalysis, as it is sometimes called) is also routinely excluded from the courts although it is an often used tool of epidemiologists.¹³⁷ Normally, several studies are

¹³¹ *Id.* at 37.

¹³² See, e.g., *Daubert*, 43 F.3d at 1320-21 (excluding causation evidence that Bendectin was capable of causing defects because expert could not testify that relative risk was greater than two).

¹³³ See Fienberg et al., *supra* note 41, at 9 (noting that "requirement of a relative risk of two . . . rests on a misinterpretation of statistical methodology").

¹³⁴ See *id.* at 9 n.27.

¹³⁵ See Sander Greenland, *Preface*, in *EVOLUTION OF EPIDEMIOLOGIC IDEAS: ANNOTATED READINGS ON CONCEPTS AND METHODS* 14 (Sander Greenland ed. 1987). Epidemiologists often refer to Hill's criteria in assessing causation. See Austin Bradford Hill, *The Environment and Disease: Association or Causation?*, 58 *PROC. ROYAL SOC'Y MED.* 295, 299 (1965) (setting out nine aspects of statistical association between two variables that ought to be considered in assessing causation). Relative risk is important according to these criteria, but Hill cautioned against dismissing a causal hypothesis merely because the observed association appears to be slight. *Id.*

¹³⁶ H. KATO, *CANCER MORTALITY IN ATOMIC BOMB SURVIVORS* (I. Shigematsu & A. Kagan, eds. 1986).

¹³⁷ The plaintiffs' experts in *Daubert* used meta-analyses as the foundation for their conclusions, and the Ninth Circuit, on remand, ultimately found it to lack scientific validity. *Daubert v. Merrell Dow Pharms., Inc.*, 43 F.3d 1311, 1312, 1320-22 (9th Cir. 1995) (explaining that in order to correct methodological errors, experts would have to change conclusions altogether).

required to reach any conclusion about the effects of a given chemical. Meta-analysis, which averages the results of many individual studies, is a way of reducing sampling error¹³⁸ by increasing the size of the sample.¹³⁹ Meta-analysis is often used to identify subtle effects of drugs in clinical studies.¹⁴⁰ Statistical reanalyses of data are also widely accepted tools in epidemiology.¹⁴¹ This technique offers a means of comparing and synthesizing studies dealing with similar health effects and risk factors.¹⁴² The appropriate use of meta-analysis is to "enhance understanding of associations" between chemical exposure and its effects on people.¹⁴³ Even if properly done, however, meta-analysis is not appropriate if there are too few studies available or when the available studies yield insufficient information about disease and exposure.¹⁴⁴ Assuming that the technique is properly implemented, one major difference between meta-analysis and other scientific methods is that in a properly performed meta-analysis, the underlying assumptions and model choices are explicit and identifiable, whereas in the latter the choices are often implicit and intuitive.¹⁴⁵ Thus, the courts' wholesale rejection of meta-analysis has little scientific justification.

4. *Conflation of Admissibility with Sufficiency.* In a toxic tort case, the plaintiff must prove causation by a preponderance of the

¹³⁸ Sampling error, sometimes called random error, refers to the difference between the estimate of a particular characteristic in a population and the true value. David H. Kaye & David A. Freeman, *Reference Guide on Statistics*, in FEDERAL JUDICIAL CTR., *supra* note 8, at 331, 407.

¹³⁹ See FOSTER & HUBER, *supra* note 40, at 80 (discussing use of meta-analyses in epidemiological studies).

¹⁴⁰ *Id.*

¹⁴¹ See, e.g., Proposed Guidelines, *supra* note 11, at 17,974 ("Statistical re-analyses of data, particularly an examination of different exposure indices, can give insight on potential exposure-response relationships."); A. Blair et al., *Guidelines for Application of Meta-Analysis in Environmental Epidemiology*, 22 REG. TOXICOLOGY PHARMACOLOGY 189, 189-97 (1995); J. Peto, *Meta-analysis of Epidemiological Studies of Carcinogenesis in Risk Assessment*, 116 IARC SCI. PUBLS. (1992).

¹⁴² Proposed Guidelines, *supra* note 11, at 17,974.

¹⁴³ *Id.*

¹⁴⁴ *Id.*

¹⁴⁵ Fienberg et al., *supra* note 41, at 14.

admissible evidence.¹⁴⁶ This is only one of several legal standards that comes into play when assessing causation evidence. The proponent must also persuade the judge that the proffered evidence is more likely than not scientifically valid.¹⁴⁷ Then there is the *Daubert* standard, which requires only a scintilla of scientifically valid and relevant evidence in order to survive an admissibility determination.¹⁴⁸ How these legal standards ought to be interpreted is far from clear and is an argument that predates *Daubert* by a considerable margin.¹⁴⁹ *Daubert* and *Joiner* have, however, muddied the waters.

In the traditional tort case, proving causation means the factfinder must be more than fifty percent confident that, having heard both sides of the story—including the scientific evidence supporting each side—the defendant's act or product caused the plaintiff's harm.¹⁵⁰ The standard does not require, however, that judges first decide whose evidence is most persuasive and exclude evidence pointing the other way.¹⁵¹ That is the jury's function. Nor should the court in an admissibility determination weigh the gravamen of the parties' proffered evidence against each other to determine whether a reasonable jury could find for the nonmoving party. That is the question of sufficiency.¹⁵² Yet, this mistaken

¹⁴⁶ See MCCORMICK ON EVIDENCE § 337 (3d ed. 1984); Green, *supra* note 100, at 680-81 ("plaintiffs should be required to prove causation by a preponderance of the available evidence, not by some predetermined standard that may require nonexistent studies").

¹⁴⁷ See Michael Hoening, *Expert Testimony by Physicians*, N.Y. L.J., Sept. 14, 1998, at 3.

¹⁴⁸ See *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 596 (1993).

¹⁴⁹ At least since Hart & Honore's critique of Hume's causal chain analysis, legal scholars have debated the meaning of causation in tort law. See H.L.A. HART & A.M. HONORE, CAUSATION IN THE LAW 80 (1959).

¹⁵⁰ See D.H. Kaye, *Apples and Oranges: Confidence Coefficients and the Burden of Persuasion*, 73 CORNELL L. REV. 54, 55 (1987) (noting that "in principle, a verdict for plaintiff is justified if an idealized judge or jury, given the parties' evidence, finds that the probability that the plaintiff's story is true exceeds some threshold figure").

¹⁵¹ See, e.g., *National Bank of Commerce v. Dow Chem. Co.*, 965 F. Supp. 1490 (E.D. Ark. 1996) (excluding scientific evidence in context of plaintiff's suit against pesticide manufacturer); see also Steve Gold, Note, *Causation in Toxic Torts: Burdens of Proof, Standards of Persuasion, and Statistical Evidence*, 96 YALE L.J. 376, 377 (1986) (observing that use of statistical probabilities of causation, as necessitated by epidemiological studies, has caused courts to conflate burdens of proof and standards of persuasion).

¹⁵² As the Supreme Court explained, "there is no issue for trial unless there is sufficient evidence favoring the nonmoving party for a jury to return a verdict for that party." *Anderson v. Liberty Lobby, Inc.*, 477 U.S. 242, 249-50 (1986). "Merely colorable" or "not significantly

conflation of sufficiency with admissibility is widespread among the federal courts.¹⁵³

The issue is rendered more confusing because, in *Daubert* hearings, the causation issue frequently surfaces in the context of a dual motion challenging the admissibility of expert testimony and moving for summary judgment.¹⁵⁴ If the expert affidavits submitted to defeat the motion for summary judgment are deemed inadmissible under *Daubert*, the motion will be granted and the case dismissed.¹⁵⁵ Thus, the first inquiry must be whether the testimony is scientifically valid. Once this determination has been made, the second inquiry is whether the totality of the defendant's evidence is sufficient to establish causation. Courts, however, including the Supreme Court in *Joiner*, conflate the inquiries finding testimony inadmissible because it is insufficient to support causation. Only if the "scintilla of evidence presented . . . is insufficient to allow a reasonable juror to conclude that the position more likely than not is true" may the judge grant summary judgment or direct a

probative" evidence will not withstand summary judgment. *Id.* On the other hand, the court is not to weigh the evidence in making this determination but is to make all inferences in favor of the nonmoving party. *Id.* at 249. If, after all, there is an absence of evidence to support the nonmoving party's case, the court may grant summary judgment. *Celotex Corp. v. Catrett*, 477 U.S. 317, 325 (1986). If the plaintiff's theory is so implausible that a jury could not rationally accept it, summary judgment may be proper. *See Matsushita Elec. Indus. Co. v. Zenith Radio Corp.*, 475 U.S. 574, 587 (1986). For a thorough discussion of summary judgment standards and the increasing misuse of summary judgment in federal courts, see Theresa M. Beiner, *The Misuse of Summary Judgment in Hostile Environment Cases*, 34 WAKE FOREST L. REV. 71 (1999).

¹⁵³ *See Conde v. Velsicol Chem. Corp.*, 24 F.3d 809, 814 (6th Cir. 1994) (holding expert testimony insufficient to allow jury to find causation due to analytical gap between evidence and inferences); *In re Hanford Nuclear Reservation Litig.*, No. CY-91-3015-AAM, 1998 WL 775340, at *140-41 (E.D. Wash. 1998) (finding unacceptable analytical gap between animal studies and proffered inferences); *National Bank of Commerce v. Dow Chem. Co.*, 965 F. Supp. 1490, 1530 (E.D. Ark. 1996) (granting defendant's motion to exclude expert testimony since plaintiffs failed to establish that "proffered expert's causation evidence meets scientific admissibility standards"); *Hall v. Baxter Healthcare, Corp.*, 947 F. Supp. 1387, 1398 (D. Or. 1996) (excluding expert testimony which fails to state causal connection in terms of probability or certainty); Thomas J. Mack, *Scientific Testimony After Daubert: Some Early Returns From Lower Courts*, TRIAL, Aug. 1994, at 23, 28-30 (reporting holdings on admissibility issues in federal courts).

¹⁵⁴ The causation issue in *Joiner*, for example, arose on defendants' motion for summary judgment. *Joiner v. General Elec. Co.*, 864 F. Supp. 1310, 1311 (N.D. Ga. 1994), *rev'd* 78 F.3d 524 (11th Cir. 1996), *rev'd* 522 U.S. 136 (1997).

¹⁵⁵ *See id.* at 1326 (finding expert opinions inadmissible under Rule 702 and *Daubert* and therefore granting defendants' summary judgment motions).

verdict.¹⁵⁶ Nothing in the *Daubert* opinion warrants conflating the two inquiries.¹⁵⁷

In *Daubert*, "fit" is a relevancy, not a sufficiency, problem.¹⁵⁸ In *Joiner*, however, Justice Rehnquist turned the question of "fit" into a sufficiency issue by sequentially excluding the studies relied on and then finding the resulting testimony insufficient to withstand summary judgment.¹⁵⁹ This approach is at odds both with the *Daubert* test of "fit" which required only minimal relevance (testimony about the phase of the moon could be admissible to show degree of darkness)¹⁶⁰ and with the minimal logical relevance the federal courts require in other contexts. It is also contrary to the explanation of the rule given by the Advisory Committee note for Federal Rule of Evidence 401 which warns against conflating sufficiency with admissibility.¹⁶¹ Conflating the two standards is pernicious because it precludes a cumulative impact analysis, the very kind of analysis that is the most scientifically reliable.¹⁶² *Daubert* explicitly separated the two considerations.¹⁶³ The real question for admissibility is whether each of the proffered studies is methodologically sound and contributes toward a biologically plausible theory of causation.¹⁶⁴ The question of sufficiency, on the other hand, is whether all the party's evidence taken together makes a plausible causation argument.¹⁶⁵ Sequentially excluding

¹⁵⁶ *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 596 (1993).

¹⁵⁷ See Polentz, *supra* note 32, at 1207.

¹⁵⁸ See *Daubert*, 509 U.S. at 591.

¹⁵⁹ *Joiner*, 522 U.S. at 141-43.

¹⁶⁰ See *Daubert*, 509 U.S. at 591 (requiring "valid scientific" connection).

¹⁶¹ FED. R. EVID. 401 advisory committee's note.

¹⁶² See Edward J. Imwinklereid, *Daubert Revisited: Disturbing Implications*, 22 CHAMPION 18, 25 (1998) (citing *Hall v. Baxter Healthcare Corp.*, 947 F. Supp. 1387, 1397 (D. Or. 1996) as an example); see also Richard D. Friedman, *Conditional Probative Value: Neoclassicism Without Myth*, 93 MICH. L. REV. 439, 473 (1994) (arguing that admissibility rules ought to be interpreted to incorporate principles of probabilistic inference so that if evidence A is ruled to be inadmissible as unhelpful, a litigant may proffer evidence B to gain admission of A, and the trial court may consider both A and B together).

¹⁶³ *Daubert*, 509 U.S. at 579.

¹⁶⁴ Imwinklereid, *supra* note 162, at 20.

¹⁶⁵ See *Matsushita Elec. Indus. Co. v. Zenith Radio Corp.*, 475 U.S. 574, 587 (1986). The question on summary judgment is whether, accepting the nonmoving party's inferences as true, a reasonable jury could find causation more probable than not. See Beiner, *supra* note 152, at 97 (arguing for a summary judgment standard similar to that used in sexual harassment cases). It is up to the jury to decide whose causation argument is more plausible.

bits of evidence as individually unable to support causation as a whole is unjustified both intellectually and procedurally.¹⁶⁶ It is a travesty of intellectual due process.

III. FEDERAL JUDICIAL CENTER GUIDELINES

Perhaps foreseeing the difficulties post-*Daubert* courts would have making the required scientific validity determinations, the Federal Judicial Center crafted its own guidelines.¹⁶⁷ It was intended as an introduction for judges who are untrained in the various fields that are frequently the source of expert testimony, and it contains basic guides to epidemiology, toxicology, survey research, DNA evidence, and statistics. The guides explicitly do not purport to instruct judges on the admissibility of various types of evidence.¹⁶⁸ Instead, they provide a primer on the methods and reasoning of these various disciplines and suggest questions that enable judges to identify disputed issues.¹⁶⁹ Each of the guides is written, at least in part, by scientists from a particular discipline.¹⁷⁰

Useful as it is in providing concise guidance on principles and methodologies, the *Reference Manual* offers no unifying theme. It gives no perspective on the scientific enterprise, instead presenting science as a series of discrete studies.¹⁷¹ Even an astute reader would be hard-pressed to discern a common thread. Unfortunately, presenting judges with fragments of the methodological details of particular disciplines has limited utility.¹⁷² Certainly, in designing

Id.

¹⁶⁶ See Beiner, *supra* note 152, at 81 (noting that the Supreme Court rejected the piecemeal exclusion of hostile work environment evidence in sexual harassment cases in favor of a "totality of the circumstances" analysis).

¹⁶⁷ FEDERAL JUDICIAL CTR., *supra* note 8.

¹⁶⁸ *Id.* at 3. These guides "are intended to assist judges in identifying the issues most commonly in dispute in these selected areas and in reaching an informed and reasoned assessment concerning the basis of expert evidence." *Id.*

¹⁶⁹ *Id.* at 3.

¹⁷⁰ For example, the *Reference Guide on Toxicology* has two authors, a toxicologist and a lawyer. Goldstein & Henifin, *supra* note 39, at 181.

¹⁷¹ See Conley & Peterson, *supra* note 9, at 1189 (observing that "science is not a conglomeration of discrete 'studies' but rather a coherent approach to analyzing the world").

¹⁷² See GROSS, *supra* note 28, at 4 (observing that "[s]cientific knowledge consists of the current answers to three questions, answers that are the product of professional conversation: What range of 'brute facts' is worth investigating? How is this range to be investigated?

experiments, scientists should avoid methodological (and logical) fallacies. It is helpful to identify some of the problems. Judges, however, are not in the business of designing experiments, rather, they are in the business of assessing the validity of imperfect experiments presented to them as evidence.¹⁷³ This problem is not addressed.

Although the *Reference Manual* does illustrate some of the major flaws that scientists attempt to avoid in designing a study, it is silent on the methodologies employed by scientists to assess completed studies. The *Reference Manual* does not explain that imperfections are unavoidable, and it does not illuminate the relative importance of various methodological failures. What is a minor flaw, and what flaw renders a study useless? The *Reference Manual* simply does not say. Thus, despite their usefulness in understanding the concerns of particular disciplines, the guidelines offer little assistance on this issue. This is unfortunate since principles of general application help to clarify the process of assessing scientific validity.¹⁷⁴ The *Reference Manual* does not assist decisionmakers in understanding the nuanced process of evaluating imperfect information.¹⁷⁵ Judges and lawyers need a more interdisciplinary approach, emphasizing the importance of the connection between scientific topics.

What do the results of these investigations mean?").

¹⁷³ The experiments judges hear about in their courtrooms are imperfect because no experiment is ever designed or executed perfectly. Although worth striving for, perfection is an elusive goal. Judges who insist on perfect studies are demanding the impossible. Indeed, a study that is too perfect is as suspect as one that is full of flaws. *Cf. infra* notes 341-348 and accompanying text.

¹⁷⁴ See GROSS, *supra* note 28, at 7-8 (discussing "stasis theory" of rhetoric in analysis of scientific explanations).

¹⁷⁵ Identifying the common methodological flaws and fallacies one is likely to encounter in a particular discipline is useful, but all methodologies are flawed and no studies are perfect, thus the inferences drawn from particular studies are only more or less apt. The trick is determining whether a particular inference will withstand scrutiny. How flawed is too flawed? This can be assessed only in connection with other similar studies. *See infra* notes 295-304 and accompanying text. The point is not whether the judge can make the ultimate decision on the aptness of the inference but that there is a good basis for the inference.

IV. EPA GUIDELINES

In contrast to both the Supreme Court's sketchy attempts to enlighten lower courts on what makes scientific assertions well grounded and the Federal Judicial Center's *Reference Manual*, which fails to illuminate a common underlying thread, the EPA has developed guidelines for agency administrators that attempt to provide a more comprehensive and scientifically justifiable framework for analysis.¹⁷⁶ Like *Daubert* decisions, the EPA's scientific validity determinations are judicially reviewed under an abuse of discretion standard.¹⁷⁷ Just as trial courts must justify admissibility determinations to appellate courts, agencies must also justify regulatory decisions on the basis of reasonable inferences from reliable evidence.¹⁷⁸ The EPA's risk assessment model recognizes the difficulty of explaining which inferences are reasonable and which evidence is reliable, thus, its framework may offer valuable insight into the issue of scientific validity even if the model is not adaptable in toto to judicial admissibility decisions.

As an initial matter, many courts reject out of hand the notion that agency decisionmaking has any relevance to judicial determinations of admissibility.¹⁷⁹ The primary reason that these courts give for their outright rejection of agency process is that it is "too speculative."¹⁸⁰ This mistaken attitude reflects the courts' arrogance and ignorance of the underlying principles and the process by which

¹⁷⁶ Proposed Guidelines, *supra* note 11.

¹⁷⁷ 5 U.S.C. § 702(2)(A) (1994) (directing courts to review agency determinations to ensure that they are not "arbitrary, capricious, an abuse of discretion or otherwise not in accordance with the law").

¹⁷⁸ See *Natural Resources Defense Council, Inc. v. EPA*, 902 F.2d 962, 968-69 (D.C. Cir. 1990) (ruling that although an agency's interpretation of equivocal facts should be given deference, a court must "carefully review the record to ascertain that the agency has made a reasoned decision based on 'reasonable extrapolations from some reliable evidence.'"), *vacated in part* 421 F.2d 326 (D.C. Cir. 1991).

¹⁷⁹ See, e.g., *Hall v. Baxter Healthcare Corp.*, 947 F. Supp. 1387 (D. Or. 1996); *Wade-Greaux v. Whitehall Lab., Inc.* 874 F. Supp. 1441 (D.V.I. 1994), *aff'd*, 46 F.3d 1120 (3d Cir. 1994).

¹⁸⁰ See, e.g., *Allen v. Pennsylvania Eng'g Corp.*, 102 F.3d 194, 196-98 (5th Cir. 1996) (rejecting plaintiff's expert witness's "weight of evidence" analysis); *Conde v. Velsicol Chem. Corp.*, 24 F.3d 809, 814 (6th Cir. 1994) (rejecting plaintiff's use of EPA draft technical support document); *In re Hanford Nuclear Reservation Litig.*, No. C4-91-3015-AAM, 1998 WL 775340, at *140-41 (E.D. Wash. 1998) (rejecting agency weight of evidence analysis).

agency decisions are made. Administrative agencies need realistic risk assessments in order to know which chemicals to regulate.¹⁸¹ If anything, agency decisions are weighed even more carefully than judicial ones because they have far-reaching effects.¹⁸² Thus, while agencies may legitimately require that the level of exposure be set low in order to guard the public health, this does not mean that they will regulate a chemical without probable cause.¹⁸³ There are powerful political and market forces acting against such unwarranted regulation.¹⁸⁴

The EPA's proposed guidelines acknowledge the importance of considering all available evidence in deciding whether a given chemical causes disease.¹⁸⁵ In order to decide whether a substance (an "agent") causes cancer (is "carcinogenic") or birth defects (is "teratogenic"), the agency examines all available evidence, including human and animal studies regarding the way a particular agent works at the cellular and subcellular levels, as well as evidence of general metabolic processes and the way that certain toxins affect the body.¹⁸⁶ The EPA refers to this kind of assessment as a "weight of the evidence" assessment.¹⁸⁷ The International Agency for

¹⁸¹ See Proposed Guidelines, *supra* note 11, at 17,965-66, 17,999.

¹⁸² See, e.g., Clayton P. Gillette & James E. Krier, *Risk, Courts, and Agencies*, 138 U. PA. L. REV. 1027, 1067 (1990) (noting restraints on agency action).

¹⁸³ *Id.*

¹⁸⁴ See FRANK P. GRAD, 4B TREATISE ON ENVIRONMENTAL LAW (1998); Troyen A. Brennan, *Environmental Torts*, 46 VAND. L. REV. 1, 21-42 (1993) (discussing political and market forces impeding regulation); Gillette & Krier, *supra* note 182, at 1058-70 (detailing anti-regulation forces).

¹⁸⁵ See Proposed Guidelines, *supra* note 11, at 17,960; see also Guidelines for Developmental Toxicity Risk Assessment, 56 Fed. Reg. 63,798, 63,816 (1991) (emphasizing importance of evaluating data from all available studies in making risk assessments).

¹⁸⁶ See Proposed Guidelines, *supra* note 11, at 17,961.

¹⁸⁷ *Id.* The agency term "weight of the evidence" should not be confused with the legal evidentiary consideration of weight which really means credibility and is determined by the fact-finder, as distinguished from an admissibility decision which the judge makes. The use of identical terminology does not signify a common meaning. The EPA's "weight of the evidence" standard simply means that the decisionmaker must weigh the cumulative impact of all available information. *Id.* In law, a judicial determination that a particular factor affects the "weight of the evidence" (rather than its admissibility) simply means that the factfinder is the proper decisionmaker. Judges frequently have difficulty deciding whether the proffered evidence is inadmissible or whether it is admissible but insufficient. See, e.g., *National Bank of Commerce v. Dow Chem. Co.*, 965 F. Supp. 1490 (E.D. Ark. 1996) (excluding evidence as insufficient).

Research on Cancer¹⁸⁸ and the National Toxicology Program¹⁸⁹ use a similar approach in evaluating scientific evidence of disease causation.¹⁹⁰ This approach acknowledges that the decisionmaker will have to assess causation based on imperfect knowledge by weighing what is known with inferences derived from general knowledge about the disease and agent.¹⁹¹

A. ANALYZING THE AVAILABLE DATA: THE FOUR PRONGS OF ASSESSMENT

The EPA Proposed Guidelines outline a method for realistic decisionmaking in the face of imperfect knowledge.¹⁹² For this purpose, the guidelines offer a series of default assumptions that vary according to the purpose of the assessment,¹⁹³ and they outline "a combination of principles and process in the application of and departure from default assumptions."¹⁹⁴ The guidelines emphasize the importance of making assumptions explicit and in giving the rationale for a particular decision.¹⁹⁵ The process is referred to as a "weight of the evidence" assessment because it acknowledges the importance of analyzing all the available data, as opposed to

¹⁸⁸ The International Agency for Research on Cancer, part of the World Health Organization, is considered one of the premiere scientific bodies for identifying disease in humans. Bruce N. Ames, *The Causes and Prevention of Cancer: The Role of the Environment*, CA51 A.L.I.-A.B.A. 49, 58 (1995).

¹⁸⁹ See National Toxicology Program, National Institute of Env'tl. Health Science, *Review of the Criteria and the Process for Preparing the Biennial Report on Carcinogens Completed: Changes Effective Immediately*, in NATIONAL TOXICOLOGY PROGRAM UPDATE 3 (1996).

¹⁹⁰ See Proposed Guidelines, *supra* note 11, at 17,964.

¹⁹¹ *Id.*

¹⁹² *Id.* at 17,960. As one scientist explained:

All scientific work is incomplete—whether it be observational or experimental. All scientific work is liable to be upset or modified by advancing knowledge. That does not confer upon us a freedom to ignore the knowledge we already have, or to postpone the action that it appears to demand at a given time.

Hill, *supra* note 135, at 295.

¹⁹³ The guidelines recognize that in assessing public health risks, the assumptions should be conservative, and that other circumstances may require different assumptions. Proposed Guidelines, *supra* note 11, at 17,965.

¹⁹⁴ *Id.*

¹⁹⁵ See *id.* (recommending that "EPA should consider adopting principles or criteria that would give greater formality and transparency to decisions to depart from defaults").

excluding some sources of information. A weight of the evidence assessment of scientific validity should not be confused with the weight of the factual evidence at trial. The latter is a function of the factfinder, whereas the former is a function of the validity finder who, after *Daubert*, is the judge. Perhaps a better way of describing this analysis is a cumulative impact analysis, as the technique involves assessing the cumulative impact of many separate bits of information on the validity of an expert's conclusion. Although the baseline assumptions may differ for agency and judicial decisionmakers, what is important, in either instance, is that the decisionmaker identify the default assumptions and explain the choice and that the choice of default assumption makes sense. This is what I call intellectual due process.¹⁹⁶

1. *Default Assumptions.* Incomplete information is inescapable.¹⁹⁷ Scientists must, and frequently do, use studies that leave out important information or that are flawed in some way.¹⁹⁸ To assess the validity of these flawed studies as building blocks for their own uses, scientists employ a number of default assumptions.¹⁹⁹ This is the same process used by the EPA.²⁰⁰

The major default assumptions listed by the Guidelines provide ways to answer each of the following questions. First, can the presence or absence of effects in one human population predict effects in a different human group?²⁰¹ Yes, says the EPA, and most

¹⁹⁶ See generally Scott Brewer, *Scientific Expert Testimony and Intellectual Due Process*, 107 YALE L.J. 1535 (1998) (discussing decisionmaking procedures used to evaluate scientific expert testimony).

¹⁹⁷ Proposed Guidelines, *supra* note 11, at 17,964.

¹⁹⁸ See GROSS, *supra* note 28, at 47 n.6 (giving example of bending of light in a gravitational field, a crucial test of general relativity). Gross explains that although the data exhibited a scatter far wider than normal canons of proof would allow, the "elegance" of the theory convinced physicists more than any notions of falsifiability. *Id.* In other words, the interpretation of the data was a creative act that built upon the work of others and might have differed in other circumstances. *Id.* That does not mean, however, that everything is "up for grabs." What it means is that judges interpreting scientific studies need to understand the context in which these studies proceed.

¹⁹⁹ See *id.* at 47 (discussing theory of falsifiability).

²⁰⁰ See Proposed Guidelines, *supra* note 11, at 17,966 (discussing major default assumptions commonly utilized in cancer risk assessments).

²⁰¹ *Id.* at 17,966-67.

courts agree.²⁰² Second, can animal studies predict effects in exposed humans?²⁰³ Yes, says the EPA, because “nearly all of the agents known to cause cancer in humans are carcinogenic in animals”;²⁰⁴ the courts are divided on this issue.²⁰⁵ Third, do metabolic pathways relate across species?²⁰⁶ Yes, according to the EPA,²⁰⁷ but most courts never get this far in their analyses.²⁰⁸

The fourth default assumption relates to whether the way toxins affect the body can be extrapolated across species.²⁰⁹ Yes, if differences in size are accounted for, according to the EPA.²¹⁰ Moreover, in the absence of contrary data, the EPA assumes that

²⁰² *Id.* at 17,967; *see, e.g.,* *Wade-Greaux v. Whitehall Labs., Inc.*, 874 F. Supp. 1441, 1451-53 (D.V.I. 1994), *aff'd*, 46 F.3d 1120 (3d Cir. 1994). The EPA cautions, however, that “this assumption could still underestimate the response of certain sensitive human subpopulations” Proposed Guidelines, *supra* note 11, at 17,967.

²⁰³ Proposed Guidelines, *supra* note 11, at 17,967-68. The EPA asserts that the basis for this default assumption is not health conservatism but models that scale metabolic rate across animals of different size. *Id.* at 17,967. This is similar to the issue of dose-response relationships which similarly assume a linear response in the absence of contrary data. *Id.*

²⁰⁴ *Id.* This is also true of birth defects. *See* Guidelines for Developmental Toxicity Risk Assessment, 56 Fed. Reg. 63,798, 63,801 (1991) (observing that nearly all agents that are teratogens in animals are also teratogenic in humans although converse is not necessarily true, as discovered in the Thalidomide fiasco—Thalidomide was not teratogenic in rodents).

²⁰⁵ *See* *Conde v. Velsicol Chem. Corp.*, 24 F.3d 809, 814 (6th Cir. 1994) (noting that draft relying on animal studies is “not probative of medical causation by its own terms”); *see also supra* notes 93-116 and accompanying text (discussing difficulty of courts in accepting validity of testimony extrapolating evidence about human injuries from animal studies). When an animal study shows no effects, the EPA’s default assumption is that there will be no effect on humans. Proposed Guidelines, *supra* note 11, at 17,967. This is not a conservative policy EPA recognizes, since there are a number of chemicals—such as arsenic—which are toxic to humans but not to animals. *Id.*

²⁰⁶ Proposed Guidelines, *supra* note 11, at 17,968.

²⁰⁷ *Id.* This is not true with respect to birth defects, where interspecies differences in gestation times and other factors means that the effects will vary. Guidelines for Developmental Toxicity Risk Assessment, 56 Fed. Reg. at 63,801 (noting that although “there is usually at least one experimental species that mimics the types of effects seen in humans, but in other species tested, the type of developmental perturbation may be different”). This leads the EPA to assume that “the types of developmental effects seen in animal studies are not necessarily the same as those that may be produced in humans.” *Id.*

²⁰⁸ *Compare* *General Elec. Co. v. Joiner*, 522 U.S. 136, 144 (1997) (excluding animal studies but not finding that they are per se inadmissible) and *National Bank of Commerce v. Dow Chem. Co.*, 965 F. Supp. 1490, 1523 (excluding animal studies), *with In re Paoli R.R. Yard PCB Litig.*, 35 F.3d 717, 779-81 (3d Cir. 1994) (finding proffered animal studies admissible).

²⁰⁹ Proposed Guidelines, *supra* note 11, at 17,968.

²¹⁰ *Id.*

the route of exposure is not material; the courts are divided.²¹¹ Fifth, the default assumption asks whether observed dose-response relationships can be correlated to lower doses.²¹² In the absence of specific data to the contrary, the EPA assumes a linear response, an approach it considers inherently conservative;²¹³ in contrast, most courts simply reject evidence based on high dosage studies as inapplicable to low dose exposures.²¹⁴ A final default assumption is one which the EPA recognizes as problematic—and which may not be appropriate for the courts—is whether benign tumors can be correlated with disease.²¹⁵ Some benign tumors later become cancerous and some do not. In the absence of data, the EPA assumes they can be correlated.²¹⁶ Most courts never reach this issue, but as a policy matter they could correctly reject this position as public health conservative.²¹⁷ On the other hand, even development of benign tumors is evidence of disease other than cancer and therefore may have important implications for human health.

The use of default assumptions is justifiable in legal terms as a form of rebuttable presumption.²¹⁸ Although scientists recognize

²¹¹ *Id.* This is in marked contrast with the courts, which frequently insist on a similar route of exposure before finding studies admissible. See, e.g., *Joiner*, 522 U.S. at 145-46 (finding animal studies inadmissible); *Wade-Greaux v. Whitehall Labs., Inc.*, 874 F. Supp 1441, 1482-84 (D.V.I. 1994) (same), *aff'd* 46 F.3d 1120 (3d Cir. 1994).

²¹² Proposed Guidelines, *supra* note 11, at 17,968-69.

²¹³ *Id.* The EPA explains that a corollary assumption is that a high dose of an agent received over a short time is equivalent to a low dose spread over a lifetime. *Id.* This too is a conservative assumption. Nonetheless, it is a workable assumption even for judicial decisionmakers because, as the Reference Guide on Toxicology explains, the use of realistic doses in animal studies leads to "a significant loss of statistical power, thereby limiting the ability of the test to detect carcinogens or other toxic compounds." Goldstein & Henifin, *supra* note 39, at 190.

²¹⁴ See, e.g., *Joiner*, 522 U.S. at 146; *Wade-Greaux*, 874 F. Supp. at 1484.

²¹⁵ Proposed Guidelines, *supra* note 11, at 17,972.

²¹⁶ *Id.*

²¹⁷ In *Joiner*, for example, although the Supreme Court mistakenly referred to the affected mice as having "cancer," it may have been substantively correct in refusing to equate their alveogenic adenoma—a benign lung tumor—with human lung cancer. *Joiner*, 522 U.S. at 144. The Court may have correctly discarded the mouse studies but for the wrong reason—not because the cancers were of different types but because the mice did not develop cancer at all. *Id.*

²¹⁸ See Troyen A. Brennan, *Helping Courts with Toxic Torts: Some Proposals Regarding Alternative Methods for Presenting and Assessing Scientific Evidence in Common Law Courts*, 51 U. PITT. L. REV. 1, 69-70 (urging adoption of scientifically sound rebuttable presumptions by courts in toxic tort cases).

that their assumptions may not hold true in every instance, they are more often than not correct. Thus, absent information to the contrary, these assumptions should be employed. With few exceptions (notably the benign tumor extrapolation), the EPA's assumptions are scientifically sound and make good policy sense for the courts. Although regulation may require a smaller incremental risk than tort liability requires, in terms of the methods and assumptions for assessing the validity of scientific studies, agencies and courts have similar goals.

2. *Hazard Assessment.* The purpose of hazard assessment is to review and evaluate data that is pertinent to answering first whether an agent poses a hazard to humans, and second, under what circumstances the hazard may be expressed.²¹⁹ The point of this assessment is to examine the "biological story the data reveal as a whole about carcinogenic effects, mode of action, and their implications for human hazard dose-response evaluation."²²⁰ The key inquiries in such characterizations include what the agency calls the conditions of expression, that is, the route of exposure, the metabolic pathways of the agent, the kinds of toxic effects it exhibits, and a comparison of these metabolic processes in humans and animals.²²¹ In making this characterization, the EPA interprets available data on the increased incidence of disease²²² in humans and animals exposed to the agent as well as any other epidemiological studies, animal studies, and data describing the chemical structure, physical properties, and function of an agent.²²³

3. *Dose-Response Assessment.* Because nearly every chemical is harmful at some dose and harmless at very low doses, the concept of dose is extremely important in assessing toxicity.²²⁴ In order to determine the maximum tolerated dose, that is, the dose above which one would expect to find disease, the EPA tests for "[t]he highest dose that causes no more than a ten percent weight

²¹⁹ Proposed Guidelines, *supra* note 11, at 17,972.

²²⁰ *Id.*

²²¹ *Id.*

²²² *Id.* Usually tumor incidence is studied in carcinogen assessments, whereas birth defects are studied in teratology.

²²³ *Id.* at 17,977.

²²⁴ See A TEXTBOOK OF MODERN TOXICOLOGY 2 (Ernest Hodgson & Patricia E. Levi eds., 2d ed. 1997).

decrement, as compared to the appropriate control groups, and does not produce mortality, clinical signs of toxicity, or pathologic lesions . . . that would be predicted to shorten the animals' natural life span."²²⁵ The maximum tolerated dose is determined by extrapolating from short-term (usually thirty to ninety day) mammal studies.²²⁶ In order to perform this extrapolation from high dose effects observed in test animals to estimate possible low dosage effects in humans, the EPA Proposed Guidelines require an explicit dose-response analysis.²²⁷ In the absence of contradictory information, the agency uses the default assumption of a linear dose-response; that is, the more of a particular agent an animal is exposed to, the proportionately higher the incidence of disease.²²⁸ Here, the key inquiry is the relationship of the dose to the intensity of the observed response.²²⁹ The EPA's preferred models are based on general concepts of mode of action and data on the agent.²³⁰ In other words, although the best dose-response models are built on actual data about how a given chemical acts in the particular species of animal or in the human body to which it was administered,²³¹ the data are almost always incomplete.²³²

If no exposure was actually observed, as is likely in environmental exposure and in most litigated toxic tort cases, the EPA recognizes that the dose-response relationship must be extrapolated. The extrapolation usually has three aspects: from high to low doses, from animal to human responses, and from one route of exposure (i.e., skin absorption, inhalation, or ingestion) to another.²³³ When adequate data are available, the EPA suggests extrapolating the

²²⁵ *Id.* at 304 (citing Proposed Guidelines, *supra* note 11).

²²⁶ *Id.* at 473.

²²⁷ *See id.* at 17,975-77.

²²⁸ *Id.* at 17,993. In other words, no dose, no response; increased dose, increased response. The Proposed Guidelines for Carcinogen Risk Assessment acknowledge that a linear response model should not be used if there are data showing nonlinear responses (where, for example, there is a threshold below which there is no observed effect, or where there are individual differences in sensitivity, or where a combination of chemicals acts in concert—like asbestos and tobacco smoke). *Id.* In those nonlinear response situations the EPA uses a margin of exposure analysis. *Id.*

²²⁹ *Id.* at 17,992.

²³⁰ *Id.*

²³¹ *Id.*

²³² *Id.*

²³³ *Id.* at 17,992.

doses used in animal studies to equivalent human doses using information on the way a particular chemical reacts in the human body.²³⁴ When such information is lacking, the EPA recommends using a default assumption of a linear relationship between dosage and body weight.²³⁵ Thus, in order to characterize a dose-response relationship, the dose data, the response data, and the extrapolation procedures must be analyzed.²³⁶ Typically, the assessment must identify the kinds of data available and account for and explain choices made with respect to these data. It must also explain the analysis in terms of the quality and quantity of the available data. In addition, the dose-response characterization must discuss the implications of variability in human susceptibility and the applicability of results to the route of exposure, dose rate, frequency, and duration.²³⁷

4. *Risk Characterization.* The purpose of risk characterization is to integrate the assessments of hazard, dose-response, and exposure in order to make risk estimates for various scenarios.²³⁸ This assessment is used by the EPA as a basis for regulatory decisionmaking and calls for an explicit evaluation of all of the available information.²³⁹ A similar integration could--and should--be outlined in a trial court's written opinion.²⁴⁰

²³⁴ *Id.* at 17,996.

²³⁵ *Id.* at 17,997 ("the default assumption [for oral exposure] is that delivered doses are related to applied dose by a power of body weight"). When extrapolating from animals to humans, the Guidelines note that the dose-response relationship and inter-species variability in sensitivity (and among people) means that while reducing the dose should reduce the risk, dose level has no effect on variability. As a default assumption to account for inter- and intra-species variation, the Guidelines use a factor of 10. *Id.* at 17,994.

²³⁶ *Id.* at 17,998.

²³⁷ *Id.* at 17,999.

²³⁸ *Id.*

²³⁹ *Id.*

²⁴⁰ Contrary to traditional practice, where trial courts do not generally issue written opinions with respect to evidentiary rulings, the results of *Daubert* hearings are widely reported. Indeed, in order for a reviewing court to undertake an abuse of discretion review, there must be an adequate record and an adequate explanation of the trial court's decision. This standard is very similar to the required review of agency determinations which mandates explicit explanations of decisions supported by an adequate record. *Arkansas v. Oklahoma*, 503 U.S. 91, 112 (1992). This evidentiary standard is commonly referred to as the "substantial evidence on the record" standard, and it ought to be used for *Daubert* determinations.

In making its risk characterization, the EPA explicitly exhorts its administrators to avoid unrealistic estimates of risk.²⁴¹ Thus, the EPA's risk characterization process should not be lightly dismissed by judges.²⁴² The goals of the proposed guidelines' risk characterization are "transparency in . . . decisionmaking, clarity in communication, consistency in core assumptions and science policies . . . , and reasonableness."²⁴³ These goals are not inapposite for judicial decisionmaking.

B. APPLICABILITY TO JUDICIAL DECISIONMAKING

Many courts²⁴⁴ and scholars²⁴⁵ object to judges making the

²⁴¹ Proposed Guidelines, *supra* note 11, at 17,999. Although the EPA recognizes that it is appropriate for a regulatory agency to err on the side of protecting health and the environment, its policy is to make assumptions as realistic as possible. *Id.*

²⁴² In *General Electric Co. v. Joiner*, 522 U.S. 136 (1997), for example, the Supreme Court all but ignored national and international environmental agencies' risk characterizations of PCBs.

²⁴³ Proposed Guidelines, *supra* note 11, at 17,999.

²⁴⁴ See, e.g., *Allen v. Pennsylvania Eng'g Corp.*, 102 F.3d 194, 196-98 (5th Cir. 1996) (rejecting plaintiffs' experts' "weight of evidence" methodology); *Conde v. Velsicol Chem. Corp.*, 24 F.3d 809, 814 (6th Cir. 1994) (rejecting plaintiffs' reliance on EPA draft document in favor of epidemiologic studies finding little evidence of adverse health effects); *In re Hanford Nuclear Reservation Litig.*, No. CY-91-3015-AAM, 1998 WL 775340, at 140-41 (E.D. Wash. 1998) (rejecting agency weight of evidence analysis); *Wade-Greaux v. Whitehall Labs., Inc.*, 874 F. Supp. 1441, *passim* (D.V.I. 1994) (rejecting plaintiffs' experts' non-epidemiologic based evidence), *aff'd*, 46 F.3d 1120 (3d Cir. 1994). Judges frequently decline to use data that were developed for risk assessment purposes because they are concerned about the conservative default assumptions that may be employed by regulatory agencies seeking to prevent disease. See, e.g., *Conde*, 24 F.3d at 814; *Hanford Nuclear Reservation Litig.*, 1998 WL 775340 at *140-41. This concern may be unwarranted in the face of the EPA's Proposed Guidelines for Carcinogen Risk Assessment which specify that causation estimates should not be unrealistic. Proposed Guidelines, *supra* note 11, at 17,999. Moreover, this concern can be addressed by using less conservative default assumptions for litigation purposes.

²⁴⁵ See, e.g., Gerald W. Boston, *A Mass-Exposure Model of Toxic Causation: The Content of Scientific Proof and The Regulatory Experience*, 18 COLUM. J. ENVTL. L. 181 (1993) (finding that courts are correct in requiring higher level of proof of causation from plaintiffs in toxic tort suits than for regulatory agencies issuing regulatory rule); Landau & O'Riordan, *supra* note 12, at 550 (arguing against permitting use of animal toxicity studies to prove causation in litigation); cf. Mark Eliot Shere, *The Myth of Meaningful Environmental Risk Assessment*, 19 HARV. ENVTL. L. REV. 409, 431 (1995) (contending that because a weight of the evidence assessment requires judgment about validity of inferences, it is beyond the EPA's expertise). But see Carl F. Cranor, *The Normative Nature of Risk Assessment: Features and Possibilities*, 8 RISK: HEALTH, SAFETY & ENV'T. 123 (1997) (arguing that risk assessment is and ought to be acknowledged as a normative endeavor and is, therefore, an appropriate decision for the courts).

assessments referred to by the agencies as "weight of the evidence" assessments, arguing that although preventive measures can be justified on imperfect knowledge, judicial decisions should not be based on anything less than certainty.²⁴⁶ This position, however, assumes that regulatory decisions made with the goal of preventing harm to human populations can afford to ignore the costs of regulation on industry.²⁴⁷ It also overlooks the great care with which regulatory risk decisions must be made.²⁴⁸ If anything, the decision to regulate is more weighty than a judicial decision because it has far greater impact.²⁴⁹ Moreover, requiring judicial decisions to be made only on the condition of perfect knowledge requires more certainty than is either possible or legally necessary.²⁵⁰ Scientists know that insisting on certain knowledge about a given agent is unreasonable and unnecessary.²⁵¹ The most that can be expected from science is a probabilistic assessment of association between cause and effect.

Probabilistic assessments inherently employ assumptions.²⁵² Scientists understand that employing default criteria like the EPA's does not make the results less valid. On the contrary, such assumptions are inescapable. There may be arguments about which assumptions are appropriate, making it important to explicitly state which assumptions are employed. In addition to the default criteria inherent in any probabilistic assessment, the default criteria used by the EPA for assessing risk are solidly grounded in scientific process and are commonly used by toxicologists and epidemiologists

²⁴⁶ See, e.g., *Allen*, 102 F.3d at 196; *Mancuso v. Consolidated Edison Co.*, 967 F. Supp. 1437, 1448 (S.D.N.Y. 1997) (citing *In re Agent Orange Products Liability Litig.*, 597 F. Supp. 740, 781 (E.D.N.Y. 1984) (Weinstein, J.)); *Cavallo v. Star Enterprise*, 892 F. Supp. 756, 773 (E.D. Va. 1995).

²⁴⁷ See, e.g., Brennan, *supra* note 184, at 21-42 (detailing care that goes into regulatory decisionmaking process).

²⁴⁸ *Id.* at 21-42.

²⁴⁹ See, e.g., FRANK P. GRAD, TREATISE ON ENVIRONMENTAL LAW (1998).

²⁵⁰ See Cranor et al., *supra* note 105, at 17 (observing that "courts have excluded sound scientific evidence with some regularity in toxic tort cases").

²⁵¹ See, e.g., Enterline, *supra* note 72.

²⁵² See Brennan & Carter, *supra* note 70, at 47 (explaining that uncertainties inherent in scientific studies do not make them illegitimate).

in evaluating each others' work.²⁵³ These default criteria are thus an important guide for the courts.

Not only is the degree of certainty required by many courts incompatible with probabilistic assessments, but such certainty is not legally mandated.²⁵⁴ Legal causation has always been a policy issue.²⁵⁵ First year torts students learn that causation is "ultimately a hodge-podge of various policy considerations, to which different courts assign various 'weights' in various cases."²⁵⁶ There are a number of contexts outside of the toxic torts arena in which courts routinely shift the initial burden of production and, more rarely, persuasion.²⁵⁷ Warrant rules such as *res ipsa loquitur*²⁵⁸ shift the

²⁵³ See, e.g., Damian Shea, *Ecological Risk Assessment*, in A TEXTBOOK OF MODERN TOXICOLOGY, *supra* note 224, at 446.

²⁵⁴ See Green, *supra* note 100, at 687 (noting that requiring statistically significant evidence is at odds with the indifference between false negatives and false positives mandated by the preponderance standard); Vern R. Walker, *Preponderance, Probability and Warranted Factfinding*, 62 BROOK. L. REV. 1075, 1080 (1996) (remarking that "attaining 'full' or 'absolute' knowledge is generally not possible in law, and is certainly not required for factfinding").

²⁵⁵ As all law students who have encountered *Palsgraf* are aware, causation itself is a matter of policy. See Richard W. Wright, *Causation in Tort Law*, 73 CAL. L. REV. 1735, 1737-39 (1985) (noting that legal realists critique of proximate cause has not resulted in a "policy-neutral account of actual causation"). As a policy matter, the level of certainty required by many courts acts as a perverse disincentive to manufacturers in terms of producing information about the risks of their products. See, e.g., Dresser, *supra* note 104 (detailing history of manufacturer neglect of basic safety testing in breast implant cases). One attempt to circumvent the causation problem resulting from information gaps is the "failure to warn" standard of liability which requires plaintiffs to show that an injury occurred and that the injury resulted from an information failure for which the defendant was responsible. RESTATEMENT (THIRD) OF TORTS: PRODUCTS LIABILITY § 2 (1997). Despite this standard, courts still insist on an unreasonably high level of certainty to establish causation. See Aaron D. Twerski & Neil B. Cohen, *Informed Decision Making and the Law of Torts: The Myth of Justiciable Causation*, 1988 U. ILL. L. REV. 607 (criticizing *Restatement* standard because it fails to ameliorate problem of high level of proof required to show causal link between failure to warn and resulting harm).

²⁵⁶ Benjamin C. Zipursky, *Legal Malpractice and the Structure of Negligence Law*, 67 FORDHAM L. REV. 649, 649 (1998). Prosser, grandfather of torts law, treats causation as a "thin shell[] into which a variety of policy judgments could be poured." *Id.* at 689 (citing KEETON ET AL., *supra* note 73); see also Richard A. Posner, *A Theory of Negligence*, 1 J. LEGAL STUD. 29, 33 (1972) (observing that "the dominant function of the fault system is to generate rules of liability that . . . will bring about . . . the efficient level of accidents and safety"); Steven Shavell, *Uncertainty Over Causation and the Determination of Civil Liability*, 28 J.L. & ECON. 587, 596 (1985) (showing that causation policies vary depending on standard of liability employed).

²⁵⁷ The best known of these presumptions is *res ipsa loquitur* in which the burden of production—and occasionally the burden of proof—is shifted from the plaintiff to the defendant. See KEETON ET AL., *supra* note 73, §§ 39-40, at 242-62 (describing cases invoking

burdens away from the plaintiff in recognition of the plaintiff's lack of access to the evidence.²⁵⁹ Decisions about the types of evidence that are sufficient to be considered "helpful" to the jury are also policy choices.²⁶⁰

Indeed, requiring that all the information gaps be filled in order to go forward is not only a scientific impossibility, but it also places a perverse disincentive on manufacturers' investment in safety testing.²⁶¹ Because any information gained from research may be used against the manufacturer in litigation, a minimum of evidence is produced. Neither industry nor government adequately funds research on potentially toxic substances.²⁶² No toxicity data exist for nearly eighty percent of the chemicals in use.²⁶³ There simply is no

res ipsa loquitur). Courts often require higher standards of proof, such as clear and convincing evidence, when factual issues trigger protective questions for one of the parties. See Walker, *supra* note 254, at 1119 (giving as examples cases where there is "possible deprivation of individual rights, a special danger of deception, or a particular type of claim disfavored on policy grounds"). Another example is the "lost chance" tort cases, where some courts alter causation theories in order to avoid the injustice of forcing the plaintiff to prove that but for the defendant's actions, the plaintiff had a greater than 50% chance of surviving. See, e.g., Werner v. Blankfort, 42 Cal. Rptr. 2d 229, 232-39 (Cal. Ct. App. 1995) (discussing range of recent positions on causation taken by courts); Evers v. Dollinger, 471 A.2d 405, 413-15 (N.J. 1984) (employing "substantial factor" causation rather than "but for" causation).

²⁵⁸ *Res ipsa loquitur*, roughly translated, means "the thing speaks for itself" and refers to circumstantial evidence that the kind of injury sustained does not generally occur in the absence of negligence. See KEETON ET AL., *supra* note 73, § 39, at 244. The policy behind this presumption is that the evidence is more readily available to the defendant than to the plaintiff, a policy equally applicable to toxic tort cases, where the manufacturers have far greater access to information about their products than do plaintiffs.

²⁵⁹ *Id.*

²⁶⁰ Legal actors need to understand that "the interpretation of uncertainties in and around . . . scientific models has been seen as a scientific matter, for scientists alone to resolve, when actually it is a process riddled with social and political implications." Brian Wynne & Sue Mayer, *How Science Fails the Environment*, 138 NEW SCI. 32, 33 (1993) (discussing role of uncertainty in science and regulatory decisionmaking); cf. Miller & Rein, *supra* note 92, at 575 (observing that "difference between the trial and appellate courts in *Joiner* could be viewed as a dispute over whether courts or juries should evaluate the proffered expert's ability to close the analytical gap between existing scientific knowledge and causality necessary to justify relief").

²⁶¹ See Dresser et al., *supra* note 104, at 731 (noting that many toxic tort cases share history of manufacturer neglect in conducting basic safety research); cf. Brennan, *supra* note 184, at 6-7 (observing that a very "weak deterrent signal" is sent by toxic torts litigation to manufacturers, making it economically feasible to ignore costs of exposing public to harmful products).

²⁶² See Wendy E. Wagner, *Choosing Ignorance in the Manufacture of Toxic Products*, 82 CORNELL L. REV. 773, 780 (1997) (citing National Research Council Report of 1984).

²⁶³ *Id.*

incentive for manufacturers to produce evidence that may eventually be used against them in court.²⁶⁴ It is more cost effective to abstain from conducting any research that is not absolutely necessary to appease the regulators.²⁶⁵ And any positive information generated, whether or not it is based on solid research, makes valid research showing adverse effects appear more controversial than it is.²⁶⁶ Indeed, there is evidence that the tobacco industry routinely paid investigators to publish highly biased reports favorable to the industry for exactly that reason.²⁶⁷ Moreover, even the small amount of research which manufacturers do support is suspect because of the potential for bias. Scientists have expressed concern for years that manufacturer support of research may bias the results.²⁶⁸

Not only do manufacturers have little incentive to engage in research on their own (or to fund the research efforts of others), but any assumption that regulatory agencies insist on testing in order to keep harmful products off the market is simply unwarranted.²⁶⁹ The tobacco, asbestos, DES, ultra-absorbent tampon, Agent Orange, breast implant, and Bendectin litigation all involved products that either were inadequately tested or which continued to be marketed

²⁶⁴ See, e.g., Wade Roush et al., *Publishing Sensitive Data: Who Calls the Shots*, 276 SCIENCE 523 (1997) (reporting dispute between researcher funded by industry and supporting manufacturer in which researcher claimed that his negative research on outbreak of lung disease at textile mill was suppressed under confidentiality agreement signed with textile company).

²⁶⁵ Notably, Bendectin was marketed for years without being tested for safety in pregnancy. See Berger, *supra* note 96, at 2146-47 (describing Merrell's development of Bendectin).

²⁶⁶ For example, the tobacco industry paid scientists to write pro-tobacco letters, which, for a considerable time, kept the tobacco issue from being litigated. See *Scientists Wrote Pro-Tobacco Letters*, ARK. DEM. GAZETTE, Aug. 5, 1998, at 3A.

²⁶⁸ See MARC A. RODWIN, *MEDICINE, MONEY AND MORALS: PHYSICIANS' CONFLICTS OF INTEREST 1* (1993) (arguing that conflict of interest is huge problem in medicine); Andrew L. Stone, *FDA: Congress Mixes Harsh Medicine*, 269 SCIENCE 1038 (1995) (observing that "conflict of interest is a real problem" in science). There have also been cases of outright fraud as described in *United States v. Keplinger*, 776 F.2d 678, 685 (7th Cir. 1985), where toxicologists falsified product safety data by underreporting animal morbidity and mortality and by omitting negative data and conclusions.

²⁶⁹ No pre-market testing on humans is required for environmental chemicals and agents. Goldstein & Henifin, *supra* note 39, at 193. As a result, less than one percent of the approximately 75,000 chemicals used in commerce have been tested for safety. *Id.* Even fewer chemicals have been subjected to epidemiological studies. *Id.* at 194.

after testing revealed adverse outcomes.²⁷⁰ Thus, the common argument against using a weight of the evidence approach based on the differing goals of courts and agencies—the goal of regulatory agencies is to keep harmful products out of circulation, whereas the goal of the courts is to assess liability for any harm caused by products that have already passed the initial regulatory hurdle—is misplaced.²⁷¹ Harmful products are not always kept out of circulation. Once in circulation, the courts' insistence on evidence that does not exist stymies the proper functioning of the tort system.

If government regulation aims to protect public health and prevent frivolous litigation, and if the tort system seeks to force manufacturers to internalize the costs of accidents and accident prevention, then good pre-marketing research should be encouraged rather than discouraged.²⁷² Permitting toxic tort plaintiffs to demonstrate causation through a weight of the evidence approach provides manufacturers with such an incentive.²⁷³ Requiring trial courts to address each of the factors and the assumptions they rely on gives the reviewing courts an opportunity to evaluate the trial court's rationale for abuse of discretion. A weight of the evidence approach is aimed at reconstructing a causal explanation of disease from whatever sources are available. Thus, with some modifications, this is an appropriate approach for the courts.²⁷⁴

²⁷⁰ *Id.* at 194 (noting repetitive pattern of manufacturer neglect in researching product safety). For example, Bendectin was marketed in 1957 as an anti-nausea drug for pregnant women before any reproductive studies had been done. It continued to be marketed without any substantial testing until the late 1970s even though the manufacturer's experience in the 1960s with thalidomide should have put it on notice that drugs taken by pregnant women can have disastrous effects on their babies. Berger, *supra* note 96, at 2144.

²⁷¹ See *Allen v. Pennsylvania Eng'g Corp.*, 102 F.3d 194, 196-98 (5th Cir. 1996) (describing different aims of regulatory law and tort law); *National Bank of Commerce v. Associate Milk Producers, Inc.*, 22 F. Supp. 2d 942, 960 (E.D. Ark. 1998) (same); *In re Joint Eastern & Southern Dist. Asbestos Litig.*, 827 F. Supp. 1014, 1029-30 (E.D.N.Y. & S.D.N.Y. 1993) (same), *rev'd* 52 F.3d 1124 (1995).

²⁷² See Dresser et al., *supra* note 104, at 708 ("the goal should be to promote good research earlier in the product development process . . . [to] protect public health and avoid unfounded litigation more effectively").

²⁷³ See Wendy E. Wagner, *The Science Charade in Toxic Risk Regulation*, 95 COLUM. L. REV. 1613, 1687 (1995) (observing that a "rational manufacturer with fiduciary obligations to shareholders . . . is thus unlikely to undertake research voluntarily").

²⁷⁴ See Cranor et al., *supra* note 105, at 53 (discussing use of animal toxicological data in making inferences about effects of various agents on humans).

A caveat is in order. Although the EPA has much to offer courts that make decisions about scientific evidence, the EPA's Proposed Guidelines cannot be adopted in toto. First, many of the guidelines are geared to assessing the risks of cancer and birth defects, whereas courts may face a much broader array of diseases.²⁷⁵ Second, many, but by no means all, of the default assumptions employed under conditions of uncertainty are more protective of the public health than may be appropriate in courts charged with assessing the validity of the evidence. Therefore, the Guidelines need to be adapted and modified in order to be useful in judicial admissibility determinations. Nonetheless, the Proposed Guidelines offer a model for intellectual due process which can be useful to the courts and fair to litigants. Taken together with an understanding of the underlying theoretical basis for scientific decisionmaking, the Proposed Guidelines may be helpful in resolving scientific issues presented to the courts.

V. MELDING SCIENCE WITH LAW: THE CAUSATION CONUNDRUM

In one respect, scientists and jurists are thoroughly divided by a common language: the conundrum of causation. Judges insist that something either causes disease or it does not.²⁷⁶ Scientists understand the matter differently.²⁷⁷ For a scientist, it is a matter of statistical attribution.²⁷⁸ If something is statistically correlated with an increased risk of disease, it may be said to "cause" the disease.²⁷⁹ Statistical explanations of causation are fundamental to

²⁷⁵ For example, toxic tort plaintiffs have claimed damages for such injuries as blindness (*Moore v. Ashland Chem., Inc.*, 151 F.3d 269 (5th Cir. 1998); *Golod v. Hoffman La Roche*, 964 F. Supp. 841 (S.D.N.Y. 1997)); damaged immune systems (*Hall v. Baxter Healthcare Corp.*, 947 F. Supp. 1387 (D. Or. 1996)); respiratory problems (*Wright v. Willamette Indus., Inc.*, 91 F.3d 1105 (8th Cir. 1996)); and headaches (*Treadwell v. Dow United Techs.*, 970 F. Supp. 974 (M.D. Ala. 1997)); among others.

²⁷⁶ See, e.g., *Moore*, 151 F.3d. at 277-79 (excluding causation evidence relating to toluene exposure); *Wright*, 91 F.3d at 1107 (dismissing formaldehyde injury claims for plaintiff's inability to demonstrate causation); *Schmaltz v. Norfolk & Western Ry. Co.*, 878 F. Supp. 1119, 1122 (N.D. Ill. 1995) (dismissing herbicide exposure case for failure to come forward with sufficient causation evidence).

²⁷⁷ Brennan & Carter, *supra* note 70, at 47.

²⁷⁸ See WESLEY C. SALMON, *THE FOUNDATIONS OF SCIENTIFIC INFERENCE* 57 (1967) (discussing mathematical theories of probability).

²⁷⁹ *Id.* at 38-39.

modern science. Probabilistic causation means that contingencies and assumptions are inextricably bound up in theory.²⁸⁰ Causes may be connected with events, but this is because "our theories connect them, not because the world is held together by cosmic glue."²⁸¹

A. PROBABILISTIC CAUSATION

A scientist's determination of causation is not an either/or proposition but a probabilistic evaluation which depends on how well all of the available evidence fits together with the underlying theory. The probability of causation can increase or decrease depending on what evidence is available.²⁸² Although the choices about the appropriate degree of conservatism to employ in making a decision about causation may vary with the context, scientists are already very conservative since scientific analyses are oriented to rejecting false positives.²⁸³

It is important to remember that association of agent and disease does not prove causation even under the best of circumstances—even, for example, in a well-conducted epidemiological study.²⁸⁴ The results of scientific studies can only tell us how frequently exposure would be associated with disease as a matter of

²⁸⁰ See, e.g., PATRICK F. SUPPES, *THE STRUCTURE OF THEORY* 186 (1976); PATRICK F. SUPPES, *A PROBABILISTIC THEORY OF CAUSALITY* (1972).

²⁸¹ NORWOOD RUSSELL HANSON, *PATTERNS OF DISCOVERY: AN INQUIRY INTO THE CONCEPTUAL FOUNDATIONS OF SCIENCE* 64 (1958).

²⁸² See Brennan & Carter, *supra* note 70, at 38-39 (discussing development of hermeneutic analysis of science that recognizes the concept of probabilistic causation).

²⁸³ Scientists attempt, whenever possible, to control and eliminate inaccuracies. *Id.* at n.158. In addition, scientific studies explicitly attempt to minimize Type II (Beta) error, the probability of failing to reject a false hypothetical. THEODORE COLTON, *STATISTICS IN MEDICINE* 128 (1974). To accomplish this, scientists demand 95% confidence intervals and corresponding p-values of 5% before they consider their studies to have statistical significance. Gad & Weil, *supra* note 36, 37, at 222-23.

²⁸⁴ See Linda A. Bailey et al., *Reference Guide on Epidemiology, in FEDERAL JUDICIAL CTR., supra* note 8, at 121, 126 ("An association identified in an epidemiological study may or may not be causal."). With respect to the correct diagnosis of disease, one court noted, "[T]hat everyone who has eaten bread has died may tell us something about bread, but not very much." *Wilkins v. University of Houston*, 654 F.2d 388, 410 (5th Cir. 1981).

chance. Probabilities help to decide whether an association between two variables exists.²⁸⁵

Statistical significance is a statement about the frequency with which a particular finding is likely to arise by chance.²⁸⁶ Confidence limits indicate the values within which a certain percentage of all data is likely to fall.²⁸⁷ Hypothesis testing, testing the hypothesis of some effect against the null hypothesis of no effect, consists of determining if two or more groups of data differ from each other at a predetermined level of confidence.²⁸⁸ If a study shows statistical significance, the null hypothesis should be rejected.²⁸⁹ In other words, the association is unlikely to be due to chance.²⁹⁰ Lack of statistical significance means that the null hypothesis cannot be rejected.²⁹¹ It does not mean that there is no association.

Biological significance and statistical significance are not the same, and sometimes they are at odds.²⁹² Even without statistical significance, the results of a study may be biologically significant, as with a toxicology study that reveals a rare tumor type appearing in more animals than expected but in too few to be statistically significant.²⁹³ The availability of a variety of studies in different disciplines that show an association between an agent and a disease

²⁸⁵ See Mervyn Susser, *Falsification, Verification and Causal Inference in Epidemiology: Reconsiderations in the Light of Sir Karl Popper's Philosophy*, in CAUSAL INFERENCE 33, 39 (Kenneth J. Rothman ed. 1988).

²⁸⁶ *Id.*

²⁸⁷ *Id.* at 40.

²⁸⁸ Gad & Weil, *supra* note 36, 37, at 256-57 (explaining importance of hypothesis testing in toxicology research).

²⁸⁹ The null hypothesis assumes that there is no association of exposure to disease. Susser, *supra* note 285, at 39.

²⁹⁰ *Id.* at 39-40.

²⁹¹ *Id.* at 40.

²⁹² See Gad & Weil, *supra* note 36, 37, at 255-56 (setting out differences between statistical and biological significance).

²⁹³ See *id.* at 261-62 (setting out example used in text); see also *supra* notes 126-127 and accompanying text (discussing courts' requirement of statistical significance for admission of scientific studies). Clinical chemistry parameters provide examples of a situation where statistical significance may be found without any biological significance. The reason for this is that biochemical parameters are rarely independent of each other. Where damage is commonly associated with an increase in three biochemicals, for example, a statistically significant rise in just one may be biologically meaningless. See Gad & Weil, *supra* note 36, 37, at 254-61.

permits scientists to make valid causal inferences without achieving statistical significance.²⁹⁴

Thus, scientists recognize that multiple confirmatory studies in different disciplines reinforce the conclusions of a single study.²⁹⁵ Biologically-based descriptions seek to describe the critical events along the causal pathway between exposure and effect.²⁹⁶ In order to do so, all available information must be assessed. There is no scientifically justifiable basis for excluding information from any chemical, cellular, or animal source that aids in this assessment.²⁹⁷

Not only should all available information be evaluated, but excluding everything but human (epidemiologic) studies for causation is a policy rather than a scientific determination.²⁹⁸ Scientists recognize that all living organisms share a common biology.²⁹⁹ This common biology leads to marked similarities in the responsiveness of subcellular structures to toxic agents.³⁰⁰ Nearly all of the chemicals recognized to cause cancer in humans also cause cancer in animals.³⁰¹ The target organs of animals and humans are frequently the same in these studies.³⁰² Moreover, animal studies are often superior to the available epidemiologic data available because of the lack of controls endemic to epidemiologic studies, the difficulty in designing and analyzing such studies, and their costliness.³⁰³ Although animal studies require extrapolation to

²⁹⁴ Bailey et al., *supra* note 284, at 126.

²⁹⁵ Goldstein & Henifin, *supra* note 39, at 212.

²⁹⁶ Rees & Hattis, *supra* note 93, at 275, 309 (describing one strategy that may be used in dealing with uncertainty in risk assessment process).

²⁹⁷ *See id.* (observing that ideal mix of information to provide complete biological explanation "would include pharmacokinetics and pharmacodynamics of the effect and its relation to the level and duration of existing exposures as well as differences in susceptibility within the exposed population" but acknowledging that such information is rarely available).

²⁹⁸ *See* Goldstein & Henifin, *supra* note 39, at 201.

²⁹⁹ *See id.* (noting that "[a]mong mammals, more than sufficient common organ structure and function readily permits the extrapolation from one species to another in most cases"). Of course, any fundamental differences in organ structure or function between the studied species and humans need to be considered. For example, rats do not have gallbladders so they cannot be used to test the effects of chemicals on human gallbladders. *Id.*

³⁰⁰ *See id.*

³⁰¹ *See* International Agency for Research on Cancer, *supra* note 106, at 17.

³⁰² *See* Huff, *supra* note 107, at 204.

³⁰³ *See* Berger, *supra* note 96, at 2128 (arguing that courts' insistence on epidemiologic studies affects outcome since epidemiologic studies are costly and take time to design, implement, and analyze). Requiring epidemiologic studies has scientific implications because

human beings, such extrapolation is a routine part of scientific analysis.³⁰⁴

As noted above, judges have trouble accepting the uncertainties of probabilistic thinking.³⁰⁵ But these uncertainties are unavoidable: although the correlation between animal data and human effects is widely accepted in the scientific community, there will always be gaps in knowledge.³⁰⁶ The exact mechanism of carcinogenesis is still unknown.³⁰⁷ Background rates of a given disease in the general population--like cancer--make it more difficult to establish a causal linkage between exposure to a potentially disease-causing agent and a disease.³⁰⁸ Further, because empirical experiments on humans are seldom possible, for ethical or feasibility reasons,³⁰⁹ many levels of uncertainty persist in epidemiological research.³¹⁰ These uncertainties although acknowledged by

it ignores the fact that such studies are prone to design errors and interpretive disputes. This unfairly skews the toxic tort system in favor of defendants. *Id.*

³⁰⁴ Proposed Guidelines, *supra* note 11, at 17,967. Regulatory agencies typically place greater weight on animal studies than they do on epidemiologic studies because animal studies are better controlled. FOSTER & HUBER, *supra* note 40, at 133. Extrapolations must be properly performed, and all parts of a given methodology must be followed. It is important, therefore, that the expert be able to explain the basis for the extrapolation used. Assuming that the extrapolations are explained and properly performed, however, there is no reason categorically to exclude animal studies.

³⁰⁵ See *supra* notes 122-136 and accompanying text.

³⁰⁶ See SAMUEL S. EPSTEIN, *THE POLITICS OF CANCER* (1978) (arguing that the politics of research funding often determines the paucity of experimental data that would link animal studies to human carcinogenesis).

³⁰⁷ See, e.g., Chemical Carcinogens: A Review of the Science and Its Associated Principles, 50 Fed. Reg. 10,372, 10,379 (1985) (acknowledging that scientists "still lack an in-depth understanding of the mechanism and stages of cancer induction and expression").

³⁰⁸ See Green, *supra* note 100, at 647 & n.20 ("At best, epidemiology assesses the likelihood that the agent caused a specific individual disease.").

³⁰⁹ See Alvin R. Feinstein, *Scientific Standards in Epidemiologic Studies of the Menace of Daily Life*, 242 SCIENCE 1257, 1257 (1988). Dr. Feinstein further observes "the public and nonepidemiologic scientists are confronted by evidence that is peer group-approved but scientifically inadequate" and suggests that "investigators will have to focus more on the scientific quality of the evidence, and less on the statistical methods of analysis and adjustment." *Id.* For purposes of evaluating scientific evidence, however, it is crucial to know the statistical assumptions used.

³¹⁰ Scientific notions of causation encompass the concept of uncertainty. For example, scientists do not like to say that a particular chemical "causes" cancer; they say there is a causal link or a high degree of correlation. Particularly in developing scientific areas, such as toxicology, all knowledge is contingent and subject to continual revision. Indeed, according to Feyerabend, it is precisely the "active interplay of various tenaciously held views" that accounts for growth in scientific knowledge. Paul Feyerabend, *Consolations for the Specialist*,

scientists, are accepted as inevitable.³¹¹ More knowledge will not cure the dilemma. There will always be uncertainties: that is the nature of probabilistic reasoning. The uncertainties do not invalidate the studies in the minds of scientists.³¹² Judges, on the other hand, are frequently paralyzed into inaction by the same uncertainties. But waiting for certainty is like waiting for Godot³¹³—an unnecessary exercise in existential angst.

Further, no matter how persuasive epidemiological or toxicological studies may be, they do not show individual, specific causation although they might enable a probabilistic judgment about the association between a particular chemical exposure and human disease on a population level.³¹⁴ This clarifies an important consideration underlying any scientific evidence: attributing causation for a particular individual is not a scientific but a legal

in CRITICISM AND THE GROWTH OF KNOWLEDGE 197, 209 (Imre Lakatos & Alan Musgrave eds., 1970) ("it is the invention of new ideas and the attempt to secure for them a worthy place in the competition that leads to the overthrow of old and familiar paradigms"). Legal actors simply cannot sit back and wait for science to sort itself out: science is continually sorting itself out. In retrospect, anomalies—those inconvenient (though valid) experimental results—have been continually piling up around the edges of accepted paradigms. See KUHN, *supra* note 28, at 121. When enough anomalies occur, scientists are challenged to account for them. The coherence of the theories which account for anomalous results and the methodologies employed are the stuff of scientific argument. Legal actors must be able to assess coherence in order to decipher the argument. For an enlightening description of the use of toxicology evidence in tort cases, see generally Ellen K. Silbergeld, *Jurisprudence and Toxicology*, in COURTS, HEALTH SCIENCE AND THE LAW 374 (1991).

³¹¹ See KENNETH J. ROTHMAN, *MODERN EPIDEMIOLOGY* 89-97, 125 (1986) (confounding factors should be identified in data analysis). Most epidemiology studies provide data showing an effect over various exposure levels in the subjects. Ideally, confounding factors, such as age, socioeconomic factors, dietary factors, and smoking, should all be taken into account. *Id.* at 90.

³¹² Indeed, epidemiologists rely on a number of operational criteria which explicitly acknowledge the probabilistic nature of the risk which is attributable to a specific exposure. See DAVID G. KLEINBAUM ET AL., *EPIDEMIOLOGIC RESEARCH: PRINCIPLES AND QUANTITATIVE METHODS* 32-34 (1982). These operational criteria, first articulated by Sir Austin Hill, include: 1) strength of association in statistical terms; 2) dose-response effect, with a higher frequency of disease at higher doses; 3) temporality: hypothesized cause must precede disease; 4) consistency of findings with other studies; 5) biological plausibility: support of the theory from biological sciences; 6) coherence (with regard to the natural history of the disease); and 7) specificity of association: disease rare outside of the exposure. *Id.* These factors must be considered together, rather than as a checklist. *Id.*

³¹³ With apologies to SAMUEL BECKETT, *WAITING FOR GODOT* (1954).

³¹⁴ Michael Dore, *A Commentary on the Use of Epidemiological Evidence in Demonstrating Cause-in-Fact*, 7 HARV. ENV'T L. REV. 429, 431 (1983).

finding.³¹⁵ Even if epidemiologists and toxicologists are able to identify correlations between exposure to a given chemical and a disease, their summary statistical statements apply only to the group studied, not to the individual members of the group.³¹⁶ Nonetheless, an overall assessment of the probabilities enables rational decisions even under conditions of uncertainty.³¹⁷

B. BURDENS OF PROOF

Some commentators have attributed the causation conundrum to the differing burdens of proof in science and law.³¹⁸ In law, the civil standard of "more probable than not" is often characterized as a probability greater than fifty percent.³¹⁹ In science, on the other

³¹⁵ See Brennan, *supra* note 102, at 512 ("Individual attribution involves uncertainty, because the epidemiological data produce only summary statistics applicable to the sample or to the population the sample represents.").

³¹⁶ *Id.*

³¹⁷ See, e.g., KLEINBAUM, *supra* note 312, at 32-34 (discussing general criteria to assess extent to which evidence supports causal interpretation); Alfred S. Evans, *Causation and Disease: A Chronological Survey*, 108 AM. J. EPID. 249, 249-58 (1978) (discussing development of concepts of causation in infectious diseases and analogizing to legal terms).

³¹⁸ See CARNEGIE COMMISSION ON SCIENCE, TECHNOLOGY AND GOVERNMENT, SCIENCE AND TECHNOLOGY IN JUDICIAL DECISIONMAKING: CREATING OPPORTUNITIES AND MEETING CHALLENGES 28 (1993) (noting that "decisions that appear to be based on 'bad' science may actually reflect the reality that the law requires a burden of proof, or confidence level, other than the 95% confidence level that is often used by scientists to reject the possibility that chance alone accounted for the observed differences").

³¹⁹ See Richard Lempert, *The New Evidence Scholarship: Analyzing the Process of Proof*, 66 B.U. L. REV. 439, 451 (1986) ("[t]he preponderance of the evidence standard is thought to mandate a verdict for the plaintiff whenever his case as a whole is more likely than not (i.e., has a greater than .50 chance) to be true"). A number of scholars argue that this is a mischaracterization, that what is really meant is that the decisionmaker has been persuaded to believe the proponent, whatever the probability of the event at issue. See, e.g., ALVIN PLANTINGA, WARRANT: THE CURRENT DEBATE 3-5 (1993); Ronald J. Allen, *The Nature of Juridical Proof*, 13 CARDOZO L. REV. 373, 416 (1991) (advocating focus on ordinal rather than cardinal proof rules); Walker, *supra* note 254, at 1078 & n.14 (stating cardinal interpretation of preponderance standard is ill-advised). For the opposite point of view, see David H. Kaye, *Do We Need a Calculus of Weight to Understand Proof Beyond a Reasonable Doubt?*, 66 B.U. L. REV. 657, 672 (1986) (arguing for cardinal interpretation). The Supreme Court explained that the goal of the legal preponderance standard is to force the litigants to "share the risk of error in roughly equal fashion." *Herman & McLean v. Huddleston*, 459 U.S. 375, 390 (1983) (citing *Addington v. Texas*, 441 U.S. 418, 423 (1979)). Professor Walker argues that the goal is not really the equal distribution of errors, but the equal treatment of similar evidence. See Walker, *supra* note 254, at 1109 (explaining that "the errors will not be distributed equally . . . unless the proportion of plaintiff verdicts equals the proportion of meritorious cases"). Professor Walker argues that the real reason for choice of preponderance

hand, the most widely used standard is a ninety-five percent confidence interval (corresponding to a five percent level of significance or p-level).³²⁰ Both sound like probabilistic assessments. As a result, the argument goes, civil judges should not exclude scientific testimony that fails scientific validity standards since the civil legal standards are much lower.³²¹ This argument is wrong in several important respects. First, its factual premise is questionable. Most excluded scientific evidence is not excluded because it fails to meet statistical significance tests but because judges fail to appreciate the inevitably uncertain nature of scientific proof and because they refuse to assess all of the available data. In other words, they require a higher standard for validity than scientists do by excluding studies that scientists routinely include in their analyses. And even for the small subset of studies rejected for failing to meet the requisite level of statistical significance, the argument misapprehends the proper basis for admissibility: not because scientific standards are too high but because judges are insisting on standards that scientists recognize to be impossibly high.

More importantly, equating confidence intervals with burdens of persuasion is simply incoherent.³²² The goal of the scientific

as the civil standard is that courts are dependent on the parties to produce evidence, and in this institutional context, the use of a mid-range decision value for factfinding creates an incentive to produce adequate evidence. *Id.* at 1114.

³²⁰ This 95% confidence level is the criterion which study data must meet before the scientist may reject the possibility that the results are due to chance. *See* Brennan, *supra* note 218, at 24 (defining confidence interval as "the range within which a study parameter lies 95% of the time").

³²¹ *Cf.* CARLF F. CRANOR, REGULATING TOXIC SUBSTANCES: A PHILOSOPHY OF SCIENCE AND THE LAW 44-48 (1993) (stating that requiring scientific standards may pose problems in protecting public health).

³²² *See, e.g., id.* (arguing that judges are wrong to exclude evidence that fails to meet statistical significance because scientific burdens of proof are set at 95% whereas legal burdens of proof are set at 50%). Professor Cranor is by no means alone in comparing statistical significance with legal burdens of proof. Among others who have made this argument are CARNEGIE COMMISSION ON SCIENCE, TECHNOLOGY, AND GOVERNMENT, SCIENCE AND TECHNOLOGY IN JUDICIAL DECISION MAKING: CREATING OPPORTUNITIES AND MEETING CHALLENGES 28 (1993); Ronald J. Allen, *Expertise and the Daubert Decision*, 84 CRIM. L. & CRIMINOLOGY 1157 (1994); Neil B. Cohen, *Confidence in Probability: Burdens of Persuasion in a World of Imperfect Knowledge*, 60 N.Y.U.L. Rev. 385 (1985); Wayne Roth-Nelson & Kathey Verdeal, *Risk Evidence in Toxic Torts*, 2 ENVTL. L. 405, 415-16 (1996)). Kaye, *supra* note 150, at 65-66. Professor Kaye likens the use of conditional error rates in deciding whether data satisfy the preponderance of the evidence standard to "trying to find the shortest path from Oxford to Cambridge by scrutinizing a map of London." *Id.* at 66.

standard—the ninety-five percent confidence interval—is to avoid claiming an effect when there is none (a false positive).³²³ Scientists who use a ninety-five percent confidence interval are making a prediction that the results are due to something other than chance. Judges, on the other hand, are simply charged with assuring that testimony is more likely than not to be based on sound reasoning; that it does not rely on incoherent probability assignments, internal contradictions, or mathematical miscalculations, and that the underlying studies describe reality as accurately as possible.³²⁴ To say that “more likely than not” means that the probability of being true is more than fifty percent, however, is wrong. The judge is not making a prediction concerning the relative frequency of accurate outcomes in repeated litigations.³²⁵ We have no way of knowing the real probability of a cause and effect relationship.³²⁶ Thus, the standpoint of the argument is incorrect. It assumes we are looking

³²³ Statistical significance tests aim to prevent the scientist from asserting a positive effect when the effect may be due to chance. See David Ozenhoff & Leslie I. Bodin, *Truth & Consequences: Health Agency Responses to Environmental Health Problems*, 12 SCI., TECH. & HUM. VALUES 70, 73-74 (1987). Statistical significance is set by convention at a level of significance, or p-value, of .05 (which corresponds to a confidence level of 95%). Fienberg et al., *supra* note 41, at 22. “By rejecting a hypothesis only when the test is statistically significant, we have placed an upper bound, 5%, on the chance of rejecting a true hypothesis.” *Id.* Another way of explaining statistical significance is that it describes the probability that the procedure produced the observed effect by chance. *Id.* at 25. If the result is not statistically significant, it may be either because the results were due to chance or because the test lacked the power to discern a difference between the null hypothesis and the proposed effect. *Id.* at 22. Power increases with the sample size of the study and with the degree of difference from the null hypothesis, i.e., the more extreme the departure from the null hypothesis, the higher the power. *Id.* Power will, therefore, be a problem for small studies of low effects—precisely those most likely to be proffered in toxic tort cases. Thus, separate studies of small numbers of subjects may not yield statistically significant results simply because each test may lack the power to distinguish the null hypothesis of no effect from patterns of illness that are not extreme. *Id.* (using employment discrimination as a hypothetical example).

³²⁴ Walker, *supra* note 254, at 1093.

³²⁵ See *id.* at 1100 (arguing that factfinder is not indicating likelihood of truth of proposition over long run). Rather, “the closest we might come to comparing significance testing with burdens of proof would be to think about equalizing the risk of rejecting the null hypothesis (no association) when it is true with the risk of accepting the alternative hypothesis (an association of some magnitude) when it is not true.” Michael D. Green, *Regulating Toxic Substances: A Philosophy of Science and the Law*, 37 JURIMETRICS J. 205, 222 (1997) [hereinafter Green, *Regulating Toxic Substances*]. This does not get us very far, however, because we presently have no way of judging accuracy.

³²⁶ Green, *Regulating Toxic Substances*, *supra* note 325, at 221.

back with the knowledge of what is an accurate outcome when all we can know is whether the evidence points toward an increased risk of association.³²⁷

In addition, equating the legal and scientific standards erroneously assumes a common view of the strengths and weaknesses of a particular study. Scientists and judges do not have such a common viewpoint, however, because the legal process makes it difficult to assess the strengths and weaknesses of a particular study.³²⁸ As a result, evidence that is uncontroversial to scientists, like the applicability of animal studies to humans, for example, appears highly contingent in court. Because of the probabilistic nature of scientific facts, based as they are on underlying assumptions, litigation makes the data appear more controversial than they actually are, creating an impression of conflict even when there is little disagreement in practice.³²⁹ Although this lack of depth

³²⁷ Professor Green contends that equating statistical significance with the legal burden of proof is wrong because it confuses whether there is any effect at all with the magnitude of any effect that may exist. Green, *supra* note 325, at 220-221. Professor Green argues that the legal standard of proof is similar to relative risk rather than statistical significance. I argue that this too is off the mark because neither the strength of association of any particular study nor the likelihood that the study's results would have occurred had there been no association is really compatible with the legal admissibility standard that the expert's testimony is more likely than not to be scientifically valid. More goes into the validity calculus than either statistical significance or relative risk. Statistical significance can measure only the first consideration. *Id.* The second consideration—magnitude of effect—is instead measured by relative risk. *Id.* Relative risk is an epidemiological term referring to the proportion of disease in exposed versus unexposed populations. David E. Austin & S. Benson Werner, *Epidemiology for the Health Sciences* 61 (1982). While significance testing characterizes the probability that the relative risk would be the same as found in the study if the results were due to chance, a relative risk of two is the threshold for a greater than fifty percent chance that the effect was caused by the agent in question (the ratio of risk in exposed versus unexposed populations, or strength of association). Green, *supra* note 325, at 222. As Professor Green explains: "The p-value tells us the probability that the study outcome . . . would occur if the null hypothesis (no association) is, in fact, true." Green, *supra* note 325, at 222. Thus, the transliteration of the "more probable than not" standard of civil factfinding into a quantitative threshold of statistical evidence is misconceived.

³²⁸ See Joseph Sanders, *From Science to Evidence: The Testimony on Causation in the Bendectin Cases*, 46 STAN. L. REV. 1, 47 (1993) (noting problem of "one-eyed factfinders" who lack "depth perception" and tend to give all scientific evidence "equal value and relevance"). Professor Sanders observed that in the litigation over the anti-nausea drug Bendectin, factfinders "learned little about the accumulation of scientific knowledge" because the lawyers spent their time trying to undermine the credibility of the witnesses. *Id.* at 47-51; see also Berger, *supra* note 96, at 2128 (noting that "deconstructed evidence all tends to sound alike").

³²⁹ See Sanders, *supra* note 328, at 48 (citing way in which illusion of conflict was created by attorneys in Bendectin litigation).

perception may affect both parties equally before a jury, it has a disparate impact on the proponent in admissibility determinations. The proponent must, after all, convince the judge that the testimony is scientifically valid.³³⁰ Creating an impression of conflict will make the judge suspicious of evidence commonly relied on by those in the scientific community, to the disadvantage of the party seeking admissibility.

The legal and scientific standards are fundamentally different. They have different goals and different measures. One cannot justifiably argue, therefore, that evidence which fails to meet the scientific standards nonetheless should be admissible because the scientific standards are too high for preponderance determinations.

Although based on a faulty premise, Professor Cranor is right in arguing that judges should not be using statistical significance—or relative risk, for that matter—as a screening device.³³¹ On the contrary, as one scientist explained, “statistical significance testing is a mechanical process that debases measurements into the qualitative and sometimes misleading categories of ‘significant’ or ‘not significant.’”³³² Making chance the primary explanation for any set of observations without thinking about what the best explanation might be is sloppy, leading to both under- and over-inclusiveness.³³³ Using statistical significance (or relative risk) as a screen

³³⁰ See *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 591-93 (1993) (describing steps in determining scientific validity).

³³¹ For cases using statistical significance or relative risk as a screening device to exclude evidence, see, e.g., *Allen v. Pennsylvania Eng'g Corp.*, 102 F.3d 194, 198-99 (5th Cir. 1996) (rejecting studies that failed to demonstrate statistical significance); *Kelley v. American Heyer-Schulte Corp.*, 957 F. Supp. 873, 884 (W.D. Tex. 1997) (requiring statistical significance); *LeBlanc v. Merrell Dow Pharms., Inc.*, 932 F. Supp. 782, 784 (E.D. La. 1996) (requiring statistically significant epidemiologic studies); *Merrell Dow Pharms., Inc. v. Havner*, 953 S.W. 2d 706, 730 (Tex. 1997) (rejecting evidence that was not statistically significant). *But see* *Berry v. C.S.X. Transp., Inc.*, 704 So. 2d 633 (Fla. App. 1997) (rejecting argument that only statistically significant associations should serve as the basis for causation opinions); *Williams v. Hedican*, 561 N.W.2d 817 (Iowa 1997) (holding trial court over-emphasized need for statistical significance).

³³² Kenneth J. Rothman, *Significance Questing*, 105 ANNALS OF INTERNAL MEDICINE 445, 445 (1986).

³³³ See *id.* Dr. Rothman explains that the unthinking use of statistical significance as a screening device leads to both under- and over-inclusiveness, making real effects appear to be due to chance and, conversely, making events that are really due to chance appear to have a real cause and effect relationship. To demonstrate the concept of over-inclusiveness, he notes that the probability of winning over a million dollars in a state lottery twice within a

ing device for admissibility is a mistake because rigid tests of statistical significance may actually reveal less about the data than other types of data analysis.³³⁴ For example, if there are multiple studies showing a small but consistent effect, scientists tend to believe that in itself is significant.³³⁵ In that way it is under-inclusive. Conversely, statistical significance in the absence of biological plausibility results in over-inclusiveness.³³⁶

Moreover, some commentators find that the root of the causation conundrum is the "amazingly uncritical" use of conventional statistical methods which incorporate the notion of chance as an explanation for the conflict of observation with prediction.³³⁷ Instead, they argue, chance should be treated for what it really is, a poor surrogate for admitting that we may not be able to account for all the conditions that determine what we observe.³³⁸ In other words, they contend, labeling a result as due to random variation is "just a way of making ignorance sound like technical explanation."³³⁹

So the unthinking use of statistical significance (or confidence intervals, or relative risk) as a prerequisite for admissibility makes little sense.³⁴⁰ Is there another justifiable set of scientific criteria that judges can use for causation determinations? One set of

short time is small enough to reject the null hypothesis but that dual jackpots are better explained by chance. *Id.* at 446. His point is that an "algorithm for inference cannot substitute for thinking about the problem." *Id.*

³³⁴ See Steven N. Goodman & Richard Royall, *Evidence and Scientific Research*, 78 AM. J. PUB. HEALTH 1568, 1568-74 (1988) (analyzing and criticizing use of p-values as evidence); Charles Poole, *Beyond the Confidence Interval*, 77 AM. J. PUB. HEALTH 195, 195-99 (1987) (noting criticisms of both significance testing and confidence intervals); W. Douglas Thompson, *Statistical Criteria in the Interpretation of Epidemiologic Data*, 77 AM. J. PUB. HEALTH 191, 191-94 (1987) (discussing lack of consensus regarding use of statistical tools in determining role of chance).

³³⁵ See Jennie A. Frieman et al., *The Importance of Beta, The Type II Error and Sample Size in the Design and Interpretation of the Randomized Control Trial: Survey of 71 "Negative" Trials*, 299 NEW ENG. J. MED. 686, 690 (1978).

³³⁶ See Rothman, *supra* note 332, at 446 (noting that statistical significance testing can "manufacture false controversies and obscure unifying quantitative interpretations").

³³⁷ See Sander Greenland, *Probability Versus Popper: An Elaboration of the Insufficiency of Current Popperian Approaches for Epidemiologic Analysis*, in CAUSAL INFERENCE 96, 102 (Kenneth J. Rothman ed. 1988); Henri Poincaré, *Chance*, in FOUNDATIONS OF SCIENCE 395 (George Bruce Halsted transl., 1946).

³³⁸ Greenland, *supra* note 337, at 102.

³³⁹ *Id.*

³⁴⁰ See Rothman, *supra* note 332, at 446-47 (excoriating use of significance testing, p-values, or confidence intervals as a substitute for thinking about data).

criteria frequently proposed are the criteria of Austin Hill.³⁴¹ But Hill himself explained that his criteria were not to be used in this way.³⁴² Koch's postulates for inferring causation of disease by microorganisms are sometimes proposed as bases for scientific validity.³⁴³ These simplistic frameworks are not useful criteria for admissibility, however, because they, like the Federal Judicial Center's *Reference Manual* guidelines, are unachievable. Few studies could—or should—hope to meet all of the criteria. Differences arise among scientists in different disciplines, and even within the same discipline regarding the amount of evidence needed to justify causation.³⁴⁴ For example, consistency of the observed effect is a criterion most scientists would consider important, but it may be absent even where there is a strong causal link.³⁴⁵ While it might be persuasive to find that there was a consistent specific association between exposure and a particular disease, such an association is rarely observed.³⁴⁶ Dose-response curves are important, but the

³⁴¹ See Hill, *supra* note 135, at 295-300. These postulates require consideration of strength of association, consistency with other scientists' results, specificity of association, temporality, biological gradient or dose-response curve, biological plausibility, coherence, and analogy with similar evidence.

³⁴² *Id.*

³⁴³ See FOSTER & HUBER, *supra* note 40, at 28 (noting that "[m]ost scientists would agree that evidence satisfying all Koch's postulates establishes a compelling case" for causation, but explaining that they are neither necessary nor sufficient conditions for inferring causation). Koch was a bacteriologist who outlined 10 criteria for causation, including: 1) higher prevalence of disease in exposed than unexposed populations; 2) those with the disease should have had more exposure to the agent than healthy populations; 3) experiments should demonstrate increased incidence of disease in exposed over unexposed populations; 4) temporality; 5) linear dose-response curve; 6) biological plausibility; and a number of other factors relating to immune responses. See Alfred S. Evans, *Causation and Disease: The Henle-Koch Postulates Revisited*, 49 *YALE J. BIOLOGY & MED.* 175, 192 (1976) (explaining that these postulates "were not regarded as rigid criteria by Koch himself and should not be today").

³⁴⁴ For an example of the debate among scientists over which chemicals are human carcinogens, compare Bruce N. Ames, *What are the Major Carcinogens in the Etiology of Human Cancer? Environmental Pollution, Natural Carcinogens, and Causes of Human Cancer: Six Errors*, in *IMPORTANT ADVANCES IN ONCOLOGY* 237 (Vincent T. Devita, Jr. et al. eds., 1989) with Jean Marx, *Animal Carcinogen Testing Challenged*, 250 *SCIENCE* 743 (1990).

³⁴⁵ Greenland, *supra* note 337, at 16-17 (citing Hill).

³⁴⁶ *Id.* at 17. This factor is often referred to by the courts as a "signature" disease although most courts recognize that it is not a prerequisite for admissibility. *In re Joint E. & S. Dist. Asbestos Litig.*, 798 F. Supp. 925, 931 (E.D.N.Y. & S.D.N.Y. 1992) (rejecting "signature disease" as only route by which plaintiff can prove causation), *rev'd on other grounds*, 998 F.2d 343 (2d Cir. 1993); cf. *National Bank of Commerce v. Dow Chem. Co.*, 965

absence of a linear response does not necessarily destroy the causal inference, although it does indicate a more complex relationship between the agent and disease.³⁴⁷ Temporality is the one criterion that is universally recognized as a prerequisite for admissibility, but everyone agrees that temporality alone is not enough.³⁴⁸ In sum, scientists recognize that what matters most is the explanatory power of the proffered theory and how well the data support the theory.

C. RESOLVING THE CONUNDRUM

The contortions over causal attribution reflect uncertainty about the interpretation of results, not their validity.³⁴⁹ Scientific observations are explained by refutable theories.³⁵⁰ The operative question is not which data can establish causation but which theories can explain the data. Causal inference is a matter of explanation.³⁵¹ The procedural jargon of scientific argument should not blind lawyers to the fact that it is nonetheless argument. The common preconception that the scientific method consists of "systematic, controlled observation or experiment whose results lead to hypotheses, which are found valid or invalid through further work, leading to theories that are reliable because they were arrived at with initial open-mindedness and continual critical skepticism" turns the process of science on its head.³⁵² As Jerome Frank remarked, the notion that "science is a charter of certainty" is an unsophisticated and unscientific view of science which any intelli-

F. Supp. 1490, 1513-14 (E.D. Ark. 1996) (finding expert testimony failed to establish "signature disease" theory).

³⁴⁷ Greenland, *supra* note 342, at 18 (citing Hill). The Federal Judicial Center's Reference Manual agrees on this point, noting that while the presence of a linear dose-response relationship strengthens the inference of causation, its absence should not be taken to weaken it. Bailey et al., *supra* note 284, at 163.

³⁴⁸ Cranor et al., *supra* note 105, at 43. For example, the fact that most dead men are bald does not necessarily mean that baldness is fatal.

³⁴⁹ See Stephan F. Lanes, *The Logic of Causal Inference in Medicine*, in CAUSAL INFERENCE, *supra* note 337, at 59, 65 (discussing interpretation of empirical evidence).

³⁵⁰ *Id.* at 66.

³⁵¹ *Id.* at 66.

³⁵² HENRY H. BAUER, SCIENTIFIC LITERACY AND THE MYTH OF THE SCIENTIFIC METHOD 19 (1992).

gent scientist would recognize as fiction “made simply to aid in getting work done, made with complete recognition of its unreality.”³⁵³

Science, no less than law, literature or philosophy, seeks to make sense of the world. Both lawyers and scientists are concerned with the presentation of evidence, and they argue about the meaning of perceived facts.³⁵⁴ While law also seeks to explain its outcomes as a search for truth,³⁵⁵ observation and experiment are not normally considered part of the legal repertoire. This apparent divergence over “genuine testability” is misleading, however. In both science and law, what counts as factual proof is a mixture of inductive and deductive reasoning hung upon a theoretical framework.³⁵⁶ Normal science operates “by extending the knowledge of those facts that the paradigm displays as particularly revealing, by increasing the extent of the match between those facts and the paradigm’s predictions, and by further articulation of the paradigm itself.”³⁵⁷ In other words, scientific reasoning, like legal reasoning, involves the use of analogy and precedent and depends for its coherence on the

³⁵³ JEROME FRANK, *LAW AND THE MODERN MIND* 307, 311 (1970).

³⁵⁴ “The practice of the scientific method is the persistent critique of arguments, in the light of tried canons for judging the reliability of the procedures by which evidential data are obtained, and for assessing the probative force of the evidence on which conclusions are based.” ERNEST NAGEL, *THE STRUCTURE OF SCIENCE: PROBLEMS IN THE LOGIC OF SCIENTIFIC EXPLANATION* 13 (1961). In legal argument, also, canons of construction are often used as focal points for argument, though their indeterminacy is acknowledged. See, e.g., KARL N. LLEWELLYN, *THE COMMON LAW TRADITION: DECIDING APPEALS* 62-120, 522-35 (1960) (setting out leeways of precedent and thrust and parry of statutory canons). The point of the “science as process” movement in science, as with the legal realism movement in law, is that while the canons themselves are indeterminate, there are evolved “steadying factors” which provide continuity and predictability. For an explanation of this in the legal context, see *id.* at 4-5 (setting out 14 clusters of factors providing predictability in the courts and explaining that “the most vital element in reckonability and stability is the courts’ constant use, in application of doctrine, and also in choosing among the branching doctrinal possibilities, of the best sense and wisdom it can muster—but always in terms of those same traditions of the work which we have seen as ‘steadying factors’”). For readings in “science as process,” see generally SHEILA JASANOFF ET AL., *HANDBOOK OF SCIENCE AND TECHNOLOGY STUDIES* 507 (1994).

³⁵⁵ See WILLIAM TWINING, *THEORIES OF EVIDENCE: BENTHAM AND WIGMORE* 12-18 (1985) (discussing rationalist tradition in evidence scholarship and its main epistemological assumption that the purpose of adjudication is to discover an objectively knowable truth).

³⁵⁶ See, e.g., Kaye, *supra* note 319, at 315 & n.9 (“[s]cientists and lawyers alike seek facts, and they understand these facts in terms of some theory”).

³⁵⁷ KUHN, *supra* note 28, at 24.

ability to discern patterns and draw upon relationships of observed phenomena. Careful construction of the hypothesis identifies, defines, and clarifies the parameters of the experiment.³⁵⁸ Analytically, both science and law rely upon subjective judgments or assumptions at every step on the way to reaching a conclusion.³⁵⁹ Science and legal analysis are also alike in that both function in relation to the social and cultural conditions from which they emerge, and both have reasonably consensual standards for testing claims within a given paradigm.³⁶⁰

Rather than focusing on "positive" and "negative" studies, interpretation should concentrate on the theory and the methodology: what alternative explanations were and were not controlled.³⁶¹ For example, concerning the association between cigarette smoking and lung cancer, the claim of a causal relation can only be justified by examining the studies and refuting the noncausal explanations.³⁶² In the absence of a competing causal theory, the theory that cigarette smoking causes lung cancer is the best available explanation of the data.³⁶³ In sum, the studies concerning causal theories should be interpreted by describing testable competing hypotheses.³⁶⁴

Practically, this ought to mean that even studies which cannot entirely support a causation hypothesis may be used in conjunction with other studies to explain the hypothesis. The critical issue for general causation is whether, based on all the evidence presented,

³⁵⁸ See Donald W. Large & Preston Michie, *Proving that the Strength of the British Navy Depends on the Number of Old Maids In England: A Comparison of Scientific Proof with Legal Proof*, 11 ENVTL. L. 555, 564 (1981).

³⁵⁹ See generally Lee Loevinger, *Standards of Proof in Science and Law*, 32 JURIMETRICS J. 323 (1992) (identifying similarities and differences between science and law in reaching proof sufficient to support a conclusion).

³⁶⁰ One statement of this credo is that "science presupposes that there are objective methods by which reliable knowledge can be tested[.] . . . hypotheses . . . can be warranted (a) by reference to the evidence, (b) by criteria of rational coherence, and (c) by their predicted experimental consequences." Paul Kurtz, *The Growth of Antiscience*, 18 SKEPTICAL INQUIRER 255, 258 (1994).

³⁶¹ See Lanes, *supra* note 349, at 66 (discussing shifting focus of interpretation from results to methods).

³⁶² See *id.* (examining competing theories of causal connection).

³⁶³ See *id.* at 70-71 (differentiating between plausible and best explanations).

³⁶⁴ See *id.* at 72 (considering this as superior approach).

an agent is more likely than not to cause disease in humans.³⁶⁵ Certainly that is the way scientists themselves assess the studies.³⁶⁶ Yet courts, including the Supreme Court,³⁶⁷ are excluding evidence simply because a single study cannot support causation. This makes little sense from a scientific or a legal viewpoint.

VI. CONCLUSION

To put the matter into perspective, many judges are demanding that scientists adopt legal definitions of scientific terms. No scientist can say with certainty that an association is causal.³⁶⁸ Even the statement that an association is probably causal is unverifiable for the simple reason that there is no way to calculate whether a given scientific theory is true or false.³⁶⁹ Accordingly, there is no way definitively to calculate that an exposure causes a disease.³⁷⁰ The most that can be expected from scientific evidence is a probabilistic statement about the association of exposure and disease.

This probabilistic statement relies on an analysis of all the available evidence in light of a biologically coherent theory of causation. Multiple sources of evidence ensure the "greatest

³⁶⁵ See *In re Paoli R.R. Yard PCB Litig.*, 35 F.3d 717, 780 (3d Cir. 1994) (acknowledging standard employed in civil litigation is that employed by EPA for classifying carcinogens).

³⁶⁶ For example, an epidemiologist reviewing 71 epidemiological studies, none of which showed statistically significant effects, nonetheless concluded that the studies "should not be viewed as conclusive given the typically small sample size that could consistently overlook important results." Frieman et al., *supra* note 335, at 690.

³⁶⁷ See *General Elec. Co. v. Joiner*, 522 U.S. 136, 146-47 (1997) (upholding the district court's ruling that because each of the proffered studies was incapable of supporting the entire causation hypothesis, none was admissible). The plaintiff and the dissent, on the other hand, argued that the studies should be assessed in conjunction with each other. *Id.* at 152-53 (Stevens, J., concurring in part and dissenting in part).

³⁶⁸ See Stephen F. Lanes, *Error and Uncertainty in Causal Inference*, in CAUSAL INFERENCE, *supra* note 337, at 173, 182 (discussing difficulty in labeling relationship or association as causal).

³⁶⁹ *Id.* at 182-83.

³⁷⁰ See *id.* Thus, there can be no empirical support for the statement that causation is the most likely explanation for an association. "The uncertainty in causal inference is attributable to the fact that we cannot establish that an association is valid." *Id.* at 185. An unidentifiable error may exist and it may cause the observation. *Id.* The most that can be expected of strength of association, the shape of a dose-response curve and the level of statistical significance is that they affect subjective beliefs. *Id.* at 185-86.

scientific certainty.”³⁷¹ The EPA Proposed Guidelines use just such a probabilistic approach. They offer salutary and much-needed guidance for the federal judiciary. With respect to the admissibility of scientific evidence, it should be sufficient that the cumulative impact of all the proffered information makes it likely that this particular chemical (or agent) is associated with human disease.³⁷² This standard translates into the familiar “more probable than not” standard—the same standard the EPA uses in making its determinations with respect to the likelihood of an agent’s causing a disease.³⁷³

Assessing all of the available evidence in conjunction with the biological plausibility of the causation theory is necessary for scientific validity. No checklist or bright-line rule can make the assessment of scientific causation evidence easy for judges. There simply is no way around having to think critically about the evidence and to weigh the studies proffered in conjunction with each other.³⁷⁴ Thus, while the legal frameworks available to judges for guidance each offer insights, none is sufficient by itself. *Daubert* and *Joiner* emphasize the importance of judicial accountability in evidentiary decisionmaking. The Federal Judicial Center’s *Reference Manual* clarifies a number of scientific concepts. The scientific validity analysis set out in the EPA’s Proposed Guidelines offers a sound methodological approach to assessing imperfect information.

³⁷¹ Cranor et al., *supra* note 105, at 29 (emphasis omitted).

³⁷² See Brennan, *supra* note 102, at 469-92 (evaluating role of various approaches to problem); see also Walker, *supra* note 254 (explaining that frequency interpretation of more probable than not makes little sense when the issue is the “chain of causal events leading to a specific plaintiff’s liver cancer”).

³⁷³ Proposed Guidelines, *supra* note 11, at 17,961.

³⁷⁴ As a result, some commentators propose abandoning the whole concept of causation in toxic tort cases, arguing that the realities of scientific uncertainty make attribution of cause and effect too difficult a task for the legal system. See Berger, *supra* note 96, at 2143. Professor Berger proposes imposing liability instead for “failure to provide substantial information relating to risk” under a theory of absolute liability for failure to meet a standard of care in developing and disseminating information. *Id.* Rather than abandoning causation, however, judges can learn to think critically about the evidence before them. Admirably, some already do so. Judge Becker, for example, has guided the Third Circuit in admissibility determinations, consistently crafting thoughtful, warranted and fair admissibility decisions. See, e.g., *In re Paoli R.R. Yard PCB Litig.*, 35 F.3d 717 (3d Cir. 1994) (showing well-reasoned decision).

Although none of these guidelines is useful alone, by combining the underlying tenets of each with a theoretical foundation based on philosophy of science teachings, the courts could make a significant step in the right direction. A heuristic built from these insights would make sound scientific and legal sense. In examining the guidelines already available to judges and suggesting a way in which the crucial insights from each could be combined, this Article offers a first step toward such a heuristic.