

1994

Scientific Validity, Admissibility, and Mass Torts after Daubert

Joseph Sanders

Follow this and additional works at: <https://scholarship.law.umn.edu/mlr>



Part of the [Law Commons](#)

Recommended Citation

Sanders, Joseph, "Scientific Validity, Admissibility, and Mass Torts after Daubert" (1994). *Minnesota Law Review*. 1708.
<https://scholarship.law.umn.edu/mlr/1708>

This Article is brought to you for free and open access by the University of Minnesota Law School. It has been accepted for inclusion in Minnesota Law Review collection by an authorized administrator of the Scholarship Repository. For more information, please contact lenzx009@umn.edu.

Scientific Validity, Admissibility, and Mass Torts After *Daubert*

Joseph Sanders*

The death of *Frye v. United States*,¹ is no longer greatly exaggerated.² *Frye* finally met its federal court demise in 1993.³ In *Daubert v. Merrell Dow Pharmaceuticals, Inc.*,⁴ the Supreme Court did what commentators⁵ had long recommended and declared that the "*Frye* test" did not survive the adoption of the Federal Rules of Evidence.⁶ The *Frye* test had declared inadmis-

* Professor of Law, University of Houston. Work on this paper was supported by the University of Houston, Environmental Liability Law Program. I wish to thank Roger Park for encouraging me to write this article and the editors at the Minnesota Law Review for their careful editing. Special thanks are due to Linda Gordon Hestor who contributed significantly to the quality of the finished product. Space limitations precluded our plans for a longer, jointly-authored article.

1. 293 F. 1013 (D.C. Cir. 1923).

2. Commentators have announced *Frye's* death on numerous occasions. See 3 JACK B. WEINSTEIN & MARGARET A. BERGER, WEINSTEIN'S EVIDENCE ¶ 702(3), at 702-36 (1993); 22 CHARLES A. WRIGHT & KENNETH W. GRAHAM, JR., FEDERAL PRACTICE AND PROCEDURE § 5168, at 86-91 (1978). Others have argued that *Frye* did survive the Federal Rules. See 1 DAVID W. LOUISELL & CHRISTOPHER B. MUELLER, FEDERAL EVIDENCE § 105, at 818 (1977).

3. *Frye's* continued vitality in state courts remains unclear. In an early state case considering the issue, the Arizona Supreme Court sidestepped the question by stating "we are not bound by the United States Supreme Court's non-constitutional construction of the Federal Rules of Evidence when we construe the Arizona Rules of Evidence." *State v. Bible*, 858 P.2d 1152, 1183 (Ariz. 1993) (en banc). For a valuable summary of state and federal law with respect to *Frye's* status, see Roger S. Hanson, *James Alphonzo Frye is Sixty-Five Years Old; Should He Retire?*, 16 W. ST. U. L. REV. 357, 372-90 (1989).

4. 113 S. Ct. 2786 (1993).

5. See Paul C. Gianelli, *The Admissibility of Novel Scientific Evidence: Frye v. United States a Half-Century Later*, 80 COLUM. L. REV. 1197, 1245-50 (1980); Mark McCormick, *Scientific Evidence: Defining a New Approach to Admissibility*, 67 IOWA L. REV. 879 (1982).

6. In so holding, the Supreme Court stated:

The drafting history makes no mention of *Frye*, and a rigid "general acceptance" requirement would be at odds with the "liberal thrust" of the Federal Rules and their "general approach of relaxing the traditional barriers to 'opinion' testimony." . . . Given the Rules' permissive backdrop and their inclusion of a specific rule on expert testimony that does not mention "general acceptance," the assertion that the Rules

sible novel expert testimony that was not "generally accepted" as reliable in the relevant scientific community.⁷ Commentators criticized *Frye* both for its unidimensional approach⁸ and for being too malleable to be useful.⁹ In *Daubert*, the Supreme Court held that Federal Rule 702 superseded *Frye* and rejected the approach followed by a majority of courts.¹⁰ Not only does *Daubert* mark the end of a long controversy over *Frye's* viability after the Federal Rules,¹¹ it also marks the end of debate on *Frye's* merits

somehow assimilated *Frye* is unconvincing. *Frye* made "general acceptance" the exclusive test for admitting expert scientific testimony. That austere standard, absent from and incompatible with the Federal Rules of Evidence, should not be applied in federal trials.

Daubert, 113 S. Ct. at 2794.

7. The D.C. Circuit reasoned as follows:

Just when a scientific principle or discovery crosses the line between the experimental and demonstrable stages is difficult to define. Somewhere in this twilight zone the evidential force of the principle must be recognized, and while courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs.

Frye, 293 F. at 1014. Over the next fifty years, many jurisdictions adopted the *Frye* rule. See EDWARD CLEARY, MCCORMICK ON EVIDENCE 606 (3d ed. 1984).

8. See Andre A. Moenssens, *Admissibility of Scientific Evidence—An Alternative to the Frye Rule*, 25 WM. & MARY L. REV. 545, 559-64 (1984).

9. The "general acceptance" test can be strictly applied to exclude all but widely accepted, mainstream scientific principles and techniques. See, e.g., *United States v. Zeiger*, 350 F. Supp. 685, 688 (D.D.C.), *rev'd*, 475 F.2d 1280 (D.C. Cir. 1972) (per curiam). Other courts have applied the test very liberally, however, prompting some observers to conclude that these judges often do little more than rely upon the opinion of a few experts. Gianelli, *supra* note 5, at 1209-11. As Gianelli notes, identifying the appropriate field within which general acceptance must be achieved can be problematic because almost all scientific techniques have received general acceptance in some narrowly defined field. *Id.* at 1211 n.95. In addition, it is not always clear what facets of the proffered testimony or underlying methodology must be "generally accepted." See Steven J. Grossman & Christopher K. Gagne, *Science and Scientific Evidence II*, 25 CONN. L. REV. 1053, 1055-57 (1993); see also *United States v. Shorter*, 809 F.2d 54 (D.C. Cir.) (discussing expert testimony concerning compulsive gambling disorders), *cert. denied*, 484 U.S. 817 (1987).

10. *Daubert*, 113 S. Ct. 2786 (1993); see PAUL C. GIANELLI & EDWARD J. IMWINKELRIED, SCIENTIFIC EVIDENCE § 1-5, at 8-13 (1993); see also *U.S. v. Smith*, 869 F.2d 348, 350 (7th Cir. 1989) (affirming the continued use of the *Frye* text).

11. See Edward R. Becker & Aviva Orenstein, *The Federal Rules of Evidence After Sixteen Years—the Effect of "Plain Meaning" Jurisprudence, the Need for an Advisory Committee on the Rules of Evidence, and Suggestions for Selective Revision of the Rules*, 60 GEO. WASH. L. REV. 857, 876-85 (1992); Margeret A. Berger, *A Relevant Approach to Novel Scientific Evidence*, 26 JURIMETICS J. 245, 246 (1986); Bert Black, *A Unified Theory of Scientific Evidence*, 56 FORDHAM L. REVIEW 595, 625-58 (1988); Ronald L. Carlson, *Policing*

as the primary device for controlling expert scientific testimony.¹²

The timing of *Frye's* rejection is of greater interest than the event itself. The Federal Rules of Evidence have been in place for nearly 20 years and, for most of that period, the circuits have disagreed about whether the Rules incorporated *Frye*. Yet only now has the Supreme Court taken the time to resolve the issue. One important reason for the Court's recent interest is a new sense of urgency concerning the increasing use of scientific expert testimony and the role judges should play in monitoring and controlling such testimony. The emerging belief that an increase in "junk science" in the courtroom¹³ requires greater judicial vigilance in admitting expert opinion has fueled this sense of urgency.¹⁴ For example, the Judicial Conference Advisory

the Bases of Modern Expert Testimony, 39 VAND. L. REV. 577, 582 n.17 (1986); Paul C. Giannelli, *Scientific Evidence: A Proposed Amendment to Federal Rule 702*, 26 JURIMETRICS J. 260, 263-64 (1986); Hanson, *supra* note 3, at 357; Edward J. Imwinkelried, *The "Bases" of Expert Testimony: The Syllogistic Structure of Scientific Testimony*, 67 N.C. L. REV. 1, 5-7 (1988); Frederick B. Lacey, *Scientific Evidence*, 24 JURIMETRICS J. 254, 266 (1984); Frederic I. Lederer, *Resolving the Frye Dilemma—A Reliability Approach*, 26 JURIMETRICS J. 240, 240 (1986); Moenssens, *supra* note 8, at 545; James E. Storrs, *Frye v. U.S. Restructured and Revitalized: A Proposal to Amend Federal Rule 702*, 26 JURIMETRICS J. 249, 250-52 (1986). The Rules' failure to clarify the standard for admitting novel scientific evidence, once called "the greatest single oversight in the Rules," precipitated this extensive literature. Becker & Orenstein, *supra*, at 877.

12. See David Bernstein, *Out of the Fryeing Pan and into the Fire: The Expert Witness Problem in Toxic Tort Litigation*, 10 REV. LITIG. 117 (1990); John D. Borders, *Fit to be Fried: Frye v. United States and the Admissibility of Novel Scientific Evidence*, 77 KY. L.J. 849 (1989); Gianelli, *supra* note 5; Frederick A. Bechtold, Note, *The Admissibility of Scientific Evidence? DNA Print Identification*, 19 STETSON L. REV. 245 (1989); Dirk Eshleman, Note, *Different Standards and Conflicting Results: A Re-Evaluation of the Frye Test for Admitting Novel Scientific Evidence in Light of Decisions Involving Spectrographic Evidence Introduction*, 5 REV. LITIG. 327 (1986).

13. PETER HUBER, GALILEO'S REVENGE: JUNK SCIENCE IN THE COURTROOM 6 (1991) [hereinafter HUBER, GALILEO'S REVENGE]; PETER HUBER, LIABILITY: THE LEGAL REVOLUTION AND ITS CONSEQUENCES 43-44 (1988); Peter Huber, *Junk Science in the Courtroom*, 26 VAL. U. LAW REV. 723 (1992); see KUNZWEIL ET. AL, THE ENVIRONMENTAL EXPERT AFTER RUBANICK AND CHRISTOPHERSEN: BEWARE THE JABBERWOCK! 1 (1991); Jeffrey K. Sherwood, *In Re Paoli Railroad: The Third Circuit Punts to the "Coffincorner"*, TOXICS L. REP., Nov. 14, 1990, at 773, 781 (criticizing the Third Circuit's approach giving plaintiffs' experts leeway in valuing non-tangible injuries).

14. See *Chaulk v. Volkswagen of America, Inc.* 808 F.2d 639, 644 (7th Cir. 1986); *Stoleson v. United States*, 708 F.2d 1217, 1222 (7th Cir. 1983); E. Donald Elliott, *Toward Incentive-Based Procedure: Three Approaches for Regulating Scientific Evidence*, 69 B.U. L. REV. 487, 489-93 (1989); Barry M. Epstein & Marc S. Klein, *The Use and Abuse of Expert Testimony in Product Liability Ac-*

Committee on Civil Rules, intending to curtail the use of expert testimony, recently proposed a change to Federal Rule 702 that would allow expert testimony only if it is "reasonably reliable and will substantially assist the factfinder."¹⁵

These developments were not lost on the Supreme Court. After dispensing with the Frye rule, the Court outlined the trial judge's gate keeping role.¹⁶ The Court's discussion raises two primary questions addressed in this Article: what approach should courts employ in assessing the admissibility of expert scientific opinion and, given this approach, how restrictive should courts be in allowing expert opinion into evidence?

Part II reviews *Daubert's* approach to admissibility. This Article argues that the concept of scientific validity lies at the heart of the Court's approach to admissibility. By taking this

tions, 17 SETON HALL L. REV. 656, 656-59 (1987); Michael S. McCarthy, Note, "Helpful" or "Reasonably Reliable?": Analyzing the Expert Witness's Methodology Under Federal Rules of Evidence 702 and 703, 77 CORNELL L. REV. 350 (1992); see also John Langbein, *The German Advantage in Civil Procedure*, 52 U. CHI. L. REV. 823, 835 (1985) (noting that judges select and commission expert witnesses in German trials); Jack B. Weinstein, *Improving Expert Testimony*, 20 U. RICH. L. REV. 473, 482 (1986) (observing that the Federal Rules allow parties to use expert testimony "to obfuscate what would otherwise be a simple case"). But see Peter Bell, *Strict Scrutiny of Scientific Evidence: A Bad Idea Whose Time Has Come*, PRODUCT SAFETY & LIABILITY REP., Jan. 17, 1992, at 79.

15. The proposed rule would read, in relevant part:

Testimony providing scientific, technical, or other specialized information, in the form of an opinion or otherwise, may be permitted only if (1) the information is reasonably reliable and will substantially assist the trier of fact to understand the evidence or to determine a fact in issue and (2) the witness is qualified as an expert by knowledge, skill, experience, training, or education to provide such testimony.

COMM. ON RULES OF PRACTICE AND PROCEDURE OF THE JUDICIAL CONFERENCE OF THE UNITED STATES, PRELIMINARY DRAFT OF PROPOSED AMENDMENTS TO THE FEDERAL RULES OF CIVIL PROCEDURE AND FEDERAL RULES OF EVIDENCE 83 (1991) (emphasis omitted). With respect to these changes, the committee noted:

[T]he revision requires that expert testimony be "reasonably reliable" and "substantially assist" the fact-finder. The rule does not mandate a return to the strictures of *Frye v. United States*, 293 F.2d [sic] 1013 (D.C. Cir., 1923) (requiring general acceptance of the scientific premises on which the testimony is based). However, the court is called upon to reject testimony that is based upon premises lacking any significant support and acceptance within the scientific community or that otherwise would be only marginally helpful to the fact-finder.

Id. at 84. For a critique of the proposed change, see Jack B. Weinstein, *Rule 702 of the Federal Rules of Evidence is Sound; It Should Not Be Amended*, 138 F.R.D. 631 (1991). The Judicial Conference Committee has deferred action and referred the issue to the new Advisory Committee on the Federal Rules of Evidence. SUMMARY OF THE REPORT OF THE COMMITTEE ON RULES OF PRACTICE AND PROCEDURE 11-12 (1992).

16. *Daubert*, 113 S. Ct. at 2795-98.

approach, the Court has invited judges to dispense with surrogate measures of scientific validity and to investigate the issue directly.¹⁷ Even though scientific validity is central to the Court's approach, the Court failed to sufficiently develop the concept in its opinion. Part III argues that this deficiency principally stems from the Court's failure to recognize that scientific validity has multiple meanings and is always a matter of degree. This Part illustrates this deficiency by sketching out four different threats to the validity of scientific research.

Although the *Daubert* Court devoted significant attention to how courts should approach the admissibility of scientific evidence, it largely failed to define how restrictive courts should be.¹⁸ The Court failed to offer any realistic examples to clarify what it means to call an expert's methodology, data, or reasoning invalid.¹⁹ Rather, the Court offered little more than the general observation that courts should judge the admissibility of scientific evidence on the basis of scientific principles of reliability and validity. In this regard, *Daubert* suffers from an ailment frequently attributed to *Frye*: it provides little guidance on how to apply Rule 703 in actual cases. The wide variation in the way the courts have applied Rules 702 and 703 in the past does little to clarify the issue.²⁰ Despite the Supreme Court's silence, how-

17. See Bert Black et. al, *Science and the Law in the Wake of Daubert: The Supreme Court Launches a New Search for Scientific Knowledge*, 72 Tex. L. Rev. (forthcoming 1994) (manuscript at 29-30, on file with author).

18. The *Daubert* Court did reject the argument that relevancy alone should govern the admissibility of scientific evidence. 113 S. Ct. at 2795.

19. The Court's unenlightening example concluded that a purported relationship between the existence of a full moon and the probability that an individual was unusually likely to have behaved irrationally would not satisfy its scientific validity standard. *Id.* at 2796.

20. See Bernstein, *supra* note 12, at 133-35; Eshleman, *supra* note 12, at 328-31. For example, the District of Columbia Circuit adopted a very passive stance in *Ferebee v. Chevron Chem. Co.*, 736 F.2d 1529 (D.C. Cir.), *cert. denied*, 469 U.S. 1062 (1984). The plaintiff claimed that exposure to paraquat, a herbicide, caused his lung disease. *Id.* at 1532. Two of the plaintiff's treating physicians proffered expert testimony on the causality issue. *Id.* at 1533. They based their opinion that paraquat exposure caused the plaintiff's disease on clinical observations of the plaintiff and the fact that one expert had identified other "similar" cases. *Id.* The court did not examine the scientific validity of this testimony, allowing the experts to testify and affirming a jury verdict for the plaintiff, noting that: "On questions such as these, which stand at the frontier of current medical and epidemiological inquiry, if experts are willing to testify that such a link exists, it is for the jury to decide whether to credit such testimony." *Id.* at 1534-39. The *Ferebee* court's description of the evidence as a "classic battle of the experts, a battle in which the jury must decide the victor," has been cited frequently. *Id.* at 1535. For a discussion of the *Ferebee* opinion, see Troyen A. Brennan, *Causal Chains and Statistical Links: The Role of Sci-*

ever, the degree to which a court may aggressively act to keep an expert's proffered testimony from the jury presents the critical issue surrounding admissibility.²¹

This Article approaches this issue in the light of scientific evidence concerning the drug Bendectin. Some of the most restrictive admissibility determinations in recent years, including *Daubert* itself, have occurred in cases involving this drug.²² Part IV reviews several Bendectin cases in which the court excluded the plaintiff's expert testimony and re-examines these rulings under a scientific validity standard. It inquires whether this standard can support the Bendectin rulings or, to put the question differently, whether the plaintiff's proffered testimony in those cases was so invalid that a court could reasonably exclude it on that basis? This Part concludes that most of the restrictive rulings are questionable under a scientific validity standard.

Despite the fact that a scientific validity standard does not justify such restrictive rulings, courts nonetheless have expressed a willingness to restrict scientific evidence in this manner. Part V provides two explanations for the judicial pro-

entific Uncertainty in Hazardous Substance Litigation, 73 CORNELL L. REV. 469, 496-97 (1988).

The District of Columbia Circuit, however, assumed a more aggressive stance in *Richardson v. Richardson-Merrell, Inc.*, affirming a trial court's grant of summary judgment for the defendant:

The question whether Bendectin causes limb reduction defects is scientific in nature, and it is to the scientific community that the law must look for the answer. For this reason, expert witnesses are indispensable in a case such as this. But that is not to say that the court's hands are inexorable tied, or that it must accept uncritically any sort of opinion espoused by an expert merely because his credentials render him qualified to testify.

857 F.2d 823, 829 (D.C. Cir. 1988), *cert. denied*, 493 U.S. 882 (1989). The court held that the expert testimony lacked an adequate foundation and was therefore inadmissible under Rule 703, distinguishing *Ferebee* on the ground that a greater body of scientific evidence existed concerning Bendectin. *Id.* at 832. It is not obvious why this distinction, even if true, should lead to the conclusion that the plaintiff's evidence lacked an adequate foundation in *Richardson* but not in *Ferebee*.

Divergent views can also be found at the trial level. Compare, for example, *Wells v. Ortho Pharmaceutical Corp.*, 615 F. Supp. 262 (N.D. Ga. 1985) (following a passive approach comparable to *Ferebee*), *aff'd*, 788 F.2d 741 (11th Cir.) (reducing damages), *cert. denied*, 479 U.S. 850 (1986), with *In re "Agent Orange" Products Liability Litigation*, 597 F. Supp. 740 (E.D.N.Y. 1984) (following a more restrictive approach comparable to *Richardson*), *aff'd*, 818 F.2d 145 (2d Cir. 1987).

21. Black et. al., *supra* note 17, at 49.

22. See, e.g., *Richardson*, 857 F.2d at 829 (addressing the issue of whether Bendectin causes limb reduction defects).

pensity to exclude scientific evidence: courts want to achieve judicial efficiency and to improve jury decision making by shielding juries from marginal science. The first objective is especially salient in mass tort cases or any situation in which the same expert testimony is repeatedly presented.²³ The second objective is relevant in all cases, including mass torts, that involve a complex body of scientific data. Although these are reasonable objectives, restrictive admissibility rulings should not be the primary means of achieving them. Several factors make restrictive evidentiary rulings especially ill-suited devices for controlling the flow of information to juries. First, achieving consistency across different types of cases is very difficult if not impossible. Second, and more importantly, marginal science is not the primary reason juries encounter substantial difficulty with scientific evidence. Rather, the primary source of difficulty is the way in which the legal process presents scientific evidence to the jury. Part VI recognizes that restrictive admissibility rulings do little to solve this problem and proposes some alternatives that will assist the fact finder in understanding and weighing scientific testimony in complex cases.

II. THE *DAUBERT* OPINION

Jason Daubert and Eric Schuler both suffer from limb reduction birth defects. They sued Merrell Dow, the manufacturer of Bendectin, claiming that the morning-sickness drug, which their mothers ingested during pregnancy, caused their defects.²⁴ The trial judge granted the defendant's motion for summary

23. For example, in the area of eyewitness identification, expert testimony rarely deals with the specifics of a given trial. Thus, experts proffer essentially the same testimony from case to case. See Roger Elliott, *Expert Testimony About Eyewitness Identification: A Critique*, 17 LAW & HUMAN BEHAVIOR 423, 423 (1993); Joseph Sanders, *Expert Witnesses in Eyewitness Facial Identification Cases*, 17 TEX. TECH L. REV. 1409, 1409-10 (1986). Appellate courts have increasingly affirmed the exclusion of such testimony. See, e.g., *United States v. Harris*, 995 F.2d 532, 534 (4th Cir. 1993); *United States v. Curry*, 977 F.2d 1042, 1051-52 (7th Cir. 1992), cert. denied, 113 S. Ct. 1357 (1993). But see, e.g., *United States v. Stevens*, 935 F.2d 1380, 1400 (3rd Cir. 1991) (reversing a decision excluding expert testimony); *Campbell v. People*, 814 P.2d 1, 8 (Colo. 1991) (also reversing a decision excluding expert testimony).

24. *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 727 F. Supp. 570, 571 (S.D. Cal. 1989), aff'd, 951 F.2d 1128 (9th Cir. 1991), vacated, 113 S. Ct. 2786 (1993). The Daubert and Schuller cases are among 1,700 or so cases brought against Merrell Dow claiming that Bendectin causes birth defects. See Joseph Sanders, *The Bendectin Litigation: A Case Study in the Life Cycle of Mass Torts*, 43 HASTINGS L.J. 301, 359-62 (1992) [hereinafter Sanders, *The Bendectin Litigation*].

judgment.²⁵ The trial court based its holding on several grounds. First, the court held that only epidemiological evidence is relevant to the question of whether Bendectin is a teratogen²⁶ and that the published epidemiological research contains no studies that demonstrate a statistically significant association between Bendectin and birth defects.²⁷ Moreover, the court found that the plaintiff's expert reanalyses of existing data, which purported to reveal a significant relationship, were insufficient to satisfy their burden of coming forward with statistically significant epidemiological evidence.²⁸ Thus, the court concluded that the strongest inference a jury could draw from the evidence was "that Bendectin could possibly have caused plaintiff's injuries," which was insufficient to avoid the defendant's motion for summary judgment.²⁹

On appeal, the Ninth Circuit affirmed in a two page opinion.³⁰ Basing its analysis on *Frye*, the Ninth Circuit held the plaintiff's expert testimony inadmissible because its underlying methodology diverged substantially from the procedures and techniques generally accepted in the field.³¹

The Supreme Court granted certiorari, primarily to announce *Frye's* demise.³² Noting the sharp division among the circuits as to *Frye's* continued vitality, the Court held that the Federal Rules of Evidence superseded the *Frye* test.³³ The

25. *Daubert*, 727 F. Supp. at 576. *Daubert* is but one of many Bendectin cases resolved at the summary judgment stage. See Joseph Sanders, *From Science to Evidence: the Testimony on Causation in the Bendectin Cases*, 46 STAN. L. REV. 1, 11, n.35 (1993) [hereinafter Sanders, *From Science to Evidence*].

26. *Daubert*, 727 F. Supp. at 575. The court explained that all other evidence lacks a sufficient foundation under Federal Rule of Evidence 703. *Id.* A teratogen is a substance that causes birth defects.

27. *Id.* The court was incorrect on this point. See *infra* note 114 and accompanying text (noting six studies finding a correlation between Bendectin use and injury).

28. *Daubert*, 727 F. Supp. at 575.

29. *Id.* at 576.

30. *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 951 F.2d 1128 (9th Cir. 1991), *vacated*, 113 S. Ct. 2786 (1993).

31. *Id.* at 1129-31.

32. 113 S. Ct. 320 (1992). The Court's refusal to grant certiorari in two other Bendectin cases, which also resulted in summary judgment for the defendant, reveals its purpose. See *Turpin v. Merrell Dow Pharmaceuticals, Inc.*, 113 S. Ct. 84 (1992) (denying petition for writ of certiorari); *Lee v. Richardson-Merrell, Inc.*, 113 S. Ct. 192 (1992) (same). The Ninth Circuit's exclusive reliance on *Frye* represents the primary distinction between these cases and *Daubert*.

33. *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 113 S. Ct. 2786, 2792-93 (1993).

Court grounded its analysis in the language of Federal Rule 702 which reads as follows:

If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, expertise, training, or education, may testify thereto in the form of an opinion or otherwise.³⁴

The Court noted that the text of the rule did not preserve the general acceptance standard, nor did its legislative history make any mention of *Frye*. The Court thus concluded that a rigid "general acceptance" standard would be contrary to the thrust of the Federal Rules which were intended to lower barriers to expert opinion testimony.³⁵

Daubert is an incomplete opinion. The Court granted certiorari primarily to announce *Frye's* death and the ensuing discussion of what standard should replace *Frye* is sketchy at best. Although the Court was quite explicit that Rule 702 does not incorporate *Frye*, it was far less clear about what Rule 702 does require.³⁶ The Court began by holding that Rule 702 modifies Rule 402's directive to admit all relevant evidence.³⁷ This holding rejected the argument that Rule 702 speaks only to the expert's credentials and that a court may admit all evidence consistent with Rules 401, 403, and 703 if presented by a qualified expert.³⁸ Rather, the Court held that Rule 702 requires reliability as well as relevance; evidence which is relevant but unreliable is inadmissible.³⁹ This interpretation of Rule 702, however, raises a fundamental question: What constitutes reliability? Importantly, the Court turned to science to answer that question: To be reliable, the offering party must have acquired the evidence through the "methods and procedures of science."⁴⁰

34. FED. R. EVID. 702.

35. *Daubert*, 113 S. Ct. at 2794 (citing *Beech Aircraft Corp. v. Rainey*, 488 U.S. 153, 169 (1988)).

36. *Daubert's* record offered little on which to base a discussion of what standard should succeed *Frye*. For this reason, Judge Weinstein argued that the Court erred in granting certiorari to *Daubert* and that a better choice would have been *Christophersen v. Allied Signal Corp.*, 939 F.2d 1106 (5th Cir. 1991), *cert. denied*, 112 S. Ct. 1280 (1992); see *Prod. Safety & Liability Rep. (BNA) 10* (Apr. 12, 1993).

37. *Daubert*, 113 S. Ct. at 2795.

38. See *Prod. Safety & Liab. Rep. (BNA) 12-30* (Apr. 26, 1993).

39. *Daubert*, 113 S. Ct. at 2795. The Court may have borrowed this analysis from the Advisory Committee's proposed change in the language of Rule 702. See *supra* note 15.

40. *Daubert*, 113 S. Ct. at 2795.

In this context, evidentiary reliability is very similar to "scientific validity."⁴¹ Although *Daubert* did not offer a systematic presentation of what scientists mean when they use this term, it did describe some broad parameters relevant to the validity inquiry. The Court emphasized that the 702 inquiry should be a flexible one and that the factors set forth in other opinions⁴² and the legal literature⁴³ may prove valuable in determining

41. *Id.* at 2795 n.9. Bert Black has proposed a modification of Rule 702 that would require scientific evidence to be based on "scientifically valid reasoning" in order to be admissible. Black, *supra* note 11, at 611.

42. *Daubert*, 113 S. Ct. at 2797. Specifically, the Court referred to the analysis in *United States v. Downing*, 753 F.2d 1224 (3d Cir. 1985). In *Downing*, the Third Circuit held that the admissibility of scientific testimony on the accuracy of eyewitness identification was "not automatic but conditional." 753 F.2d at 1226. In order to be admissible, evidence must survive the trial court's preliminary inquiry. In an *in limine* proceeding, the judge should balance: (1) the reliability of the scientific principles the expert employed; against (2) the likelihood that the evidence may overwhelm or mislead the jury. *Id.* In addition, the trial court should examine the "fit" between the proffered scientific testimony and the contested issues in the case. *Id.* at 1226. For a discussion of *Downing* by Judge Becker, its author, see Becker & Orenstein, *supra* note 11, at 881.

The court in *Christophersen v. Allied-Signal Corp.* set out a similar test for admissibility. 939 F.2d 1106 (5th Cir. 1991) (en banc), cert. denied, 112 S. Ct. 1280 (1992). In *Christopherson* the Fifth Circuit, sitting *en banc*, sustained the trial judge's grant of summary judgment to the defendant. *Id.* at 1116. The plaintiff had argued that exposure to nickel/cadmium caused her husband's fatal colon cancer. *Id.* at 1108. The *en banc* opinion established a four factor test of admissibility:

- (1) Whether the witness is qualified to express an expert opinion [under Rule 702];
- (2) whether the facts upon which the expert relies are the same type as are relied upon by other experts in the field[, as Rule 703 requires];
- (3) whether in reaching his conclusion the expert used a well-founded methodology [under *Frye*]; and
- (4) assuming the expert's testimony has passed Rules 702, 703, and the *Frye* test, whether . . . the testimony's potential for unfair prejudice substantially outweighs its probative value [under Rule 403].

Id. at 1110. The court noted that these four factors "lend themselves to sequential application." *Id.* For a further discussion of the *Christopherson* case, see Bruce James, *Frye Expert Witnesses: The 5th Circuit Takes Charge of Scientific Testimony*, 12 REV. LITIG. 171, 188 (1992).

43. *Daubert*, 113 S. Ct. at 2797 n.12. The court cites 3 WEINSTEIN & BERGER, WEINSTEIN'S EVIDENCE ¶ 702[03], at 702-41, 702-42 and Mark McCormick, *Scientific Evidence: Defining a New Approach to Admissibility*, 67 IOWA L. REV. 879, 911-912 (1982), both of which appear to have been taken from Black, *supra* note 11, at 642, n.258. Weinstein & Berger list seven factors that a court may use in assessing scientific evidence: (1) the technique's general acceptance in the field; (2) the expert's qualifications and stature; (3) the use which has been made of the technique; (4) the potential rate of error; (5) the existence of a specialized literature; (6) the novelty of the invention; and (7) the extent to which the technique relies on the expert's subjective interpretation. WEINSTEIN &

whether scientific testimony is reliable. The trial judge should determine whether proffered evidence is scientifically valid by examining the reasoning and methodology underlying the expert's testimony and the "fit" between the testimony and the factual issue presented to the judge or jury.⁴⁴ Generally, the expert's theory must be both testable and falsifiable.⁴⁵ The unreliability of a procedure and its potential rate of error⁴⁶ may likewise merit exclusion.⁴⁷

Moreover, the trial court may consider a number of secondary, surrogate indicia of reliability. These include whether the theory or technique has been subject to peer review,⁴⁸ whether the results have been published⁴⁹ and, in a partial resurrection of the *Frye* test, whether the expert's methods and reasoning enjoy general acceptance in a relevant scientific community.⁵⁰ Unlike the *Frye* test, however, which determines the value of

BERGER, *supra* note 2, ¶ 702[03], at 702-41, 702-42, quoted in Black, *supra* note 11, at 642. Black also summarized eleven factors set forth by McCormick: (1) the technique's potential error rate; (2) the existence and maintenance of standards governing its use; (3) the presence of safeguards in the technique's characteristics; (4) analogy to other scientific techniques whose results are admissible; (5) the extent to which scientists in the relevant field have accepted the technique; (6) the nature and breadth of the inference adduced; (7) the clarity and simplicity with which the technique can be described and its results explained; (8) the extent to which the courts and jury can verify the basic data; (9) the availability of other experts to test and evaluate the technique; (10) the evidence's probative significance in the circumstances of the case; and (11) the care with which the expert employed the technique. Black, *supra* note 11, at 642 n.258 (quoting McCormick, *supra* note 5, at 911-912).

44. "An additional consideration under Rule 702—and another aspect of relevancy—is whether expert testimony proffered in the case is sufficiently tied to the facts of the case that it will aid the jury in resolving a factual dispute." *United States v. Downing*, 753 F.2d 1224, 1242 (3d Cir. 1985). In *Daubert*, the Supreme Court did not directly relate its discussion of "fit" to questions of validity. Nevertheless, it seems reasonable to apply that concept in this context. A particular piece of research may support one conclusion and not another. This is frequently a question of external validity. See *infra* part III.D.

45. *Daubert*, 113 S. Ct. at 2796.

46. *Id.* at 2796-97 (citing *United States v. Smith*, 869 F.2d 348, 353-354 (7th Cir. 1989)).

47. The Court noted, almost in passing, that the "focus, of course, must be solely on principles and methodology, not the conclusions that they generate." *Id.* at 2797. This statement will likely generate a good deal of controversy. Bendectin plaintiffs have already picked up on this point and argued that their expert testimony should not be excluded under *Daubert* when its methodology is sound and the defense only objects to the expert's conclusion. See *BENDEC-TIN: Plaintiffs seek rehearing after CA 3 Affirms Defense Judgment*, *Prod. Safety & Liab. Rep. (BNA)* 2 (Sept. 1, 1993).

48. *Daubert*, 133 S. Ct. at 2797.

49. *Id.*

50. *Id.*

science primarily through the surrogate of general acceptance,⁵¹ the *Daubert* surrogates are secondary to a direct analysis of the testimony's scientific validity.

Finally, the Court also noted that Rule 702 does not stand alone. Rule 703 provides that a court may admit expert scientific opinion only if the facts or data are "of a type reasonably relied upon by experts in the particular field in forming opinions or inferences on the subject."⁵² Rule 706 allows the court to appoint its own expert when necessary.⁵³ Finally, the court may employ Rule 403 to exclude expert testimony when its prejudicial effect or potential to confuse or mislead the jury substantially outweighs its probative value.⁵⁴

In sum, *Daubert* clearly ended *Frye's* reign in the federal courts. Rules 702 and 703 superseded *Frye* and supplanted its "general acceptance" standard. Unfortunately, *Daubert* was far less clear about precisely what these rules, especially Rule 702, require. At the core of the Court's analysis of admissibility under Rule 702, however, is the idea of scientific validity. The

51. See Black, *supra* note 11, at 629.

52. *Daubert*, 113 S. Ct. at 2797. The full text of Rule 703 reads:

The facts or data in the particular case upon which an expert bases an opinion or inference may be those perceived or made known to the expert at or before the hearing. If of a type reasonably relied upon by experts in a particular field in forming opinions or inferences upon the subject, the facts or data need not be admissible in evidence.

FED. R. EVID. 703.

There has been confusion concerning the relationship of Rules 702 and 703. See, e.g., *Zenith Radio Corporation v. Matsushita Elect. Indus., Inc.*, 505 F. Supp. 1313, 1318 (E.D. Pa. 1980) (discussing the general relationship among rules 702, 703 and 704), *aff'd in part, rev'd in part*, 723 F.2d 238 (3d Cir. 1983), *rev'd*, 475 U.S. 574 (1986). Professor Imwinkelried has proposed that courts use Rule 702 to regulate the expert's major premise—the principles and theories upon which the expert bases its opinion, and that courts use Rule 703 to address the expert's minor premise—how the principles and theories apply to the facts and data in the case at hand. See Imwinkelried, *supra* note 11, at 14-16, 16-19. When the expert's testimony follows this major premise-minor premise format, Professor Imwinkelried's analysis may prove especially useful. Expert testimony, however, is not always easily broken into these two categories. See David Faigman, *Struggling to Stop the Flood of Unreliable Expert Testimony*, 76 MINN. L. REV. 877, 886 (1992).

53. See JOE S. CECIL & THOMAS WILLGING, COURT APPOINTED EXPERTS: DEFINING THE ROLE OF EXPERTS APPOINTED UNDER FEDERAL RULE OF EVIDENCE 706 (1993).

54. The full text of Rule 403 reads: "Although relevant, evidence may be excluded if its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury, or by considerations of undue delay, waste of time, or needless presentation of cumulative evidence." FED. R. EVID. 403.

next section attempts to remedy *Daubert's* failure to define this term.

III. THE VARIETIES OF SCIENTIFIC VALIDITY

Daubert's references to validity pose two primary problems: the Court used the term as if it encompassed a unitary concept with a single meaning; and the Court implied that validity must be either present or absent and not a matter of degree.⁵⁵ Validity, however, is a complex concept with multiple dimensions. For example, Cook and Campbell have identified four basic types of validity: statistical conclusion validity, internal validity, construct validity, and external validity.⁵⁶ Each may be threatened to various degrees and in a number of ways.

A. STATISTICAL CONCLUSION VALIDITY

Statistical conclusion validity is an important consideration with all quantitative data. The typical threats to statistical conclusion validity have been widely discussed.⁵⁷ When researchers observe a co-variation between two variables, they may wish to conclude, based on a statistical analysis, that the variables are causally related. Tests of significance guard against the danger that researchers will conclude that a relationship exists when it does not. Typically, tests of statistical significance test the null hypothesis that no relationship exists between two variables—that the relationship observed could be the result of chance.⁵⁸ Unless a relationship is statistically significant, the null hypothesis will not be rejected.⁵⁹ These tests, therefore, guard against Type I errors, a validity threat that occurs when one concludes that a relationship exists when, in fact, none

55. *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 113 S. Ct. 2786, 2795, n.9 (1993).

56. THOMAS D. COOK & DONALD T. CAMPBELL, *QUASI-EXPERIMENTATION: DESIGN AND ANALYSIS ISSUES FOR FIELD TESTING* 37-39 (1979). The authors do not claim that this list is exhaustive. *Id.*

57. See, e.g., Michael D. Green, *Expert Witnesses and Sufficiency of Evidence in Toxic Substances Cases: The Legacy of Agent Orange and Bendectin Litigation*, 86 Nw. U.L. REV. 643, 682-94 (1992); Sanders, *The Bendectin Litigation*, *supra* note 24, at 342-45.

58. FREDERICK WILLIAMS, *REASONING WITH STATISTICS: HOW TO READ RESEARCH* 54 (3d ed. 1986).

59. By convention, the null hypothesis will not be rejected unless the probability that chance caused a result is less than one in 20 (Alpha = .05) or, occasionally, less than one in 100 (Alpha = .01). *Id.* at 59.

does.⁶⁰ Thus, a researcher may reject a causal interpretation of an apparent relationship that is not statistically significant.

Recent interest has focused on the problem which arises when researchers attempt to assess causation with respect to rare events such as limb reduction birth defects.⁶¹ Whenever a study has a relatively small number of subjects, statistical tests will fail to detect a significant difference unless the relationship is quite strong. This may cause a Type II error, which occurs when one accepts the null hypothesis as true when it is actually false. This threat to validity is one of low statistical power.⁶² Researchers can employ a number of techniques to guard against this threat, including meta-analysis which increases the number of cases by combining the results of several studies.⁶³

Two other threats to statistical conclusion validity deserve special mention. First, there is the error rate problem. Researchers engaged in a fishing expedition, sifting through a large number of correlations in search of significant relationships, will inevitably find some. For example, if one concludes that a relationship is significant if there is less than one chance in twenty ($\text{Alpha} = .05$) that it would occur if the null hypothesis is correct, a study of sixty relationships will produce three significant correlations even if no true causal relationships exist.⁶⁴ The unreliability of measurement techniques pose the second threat. Epidemiological research depends on determining whether individuals have or have not been exposed to a toxic substance and whether or not they suffer from some adverse ef-

60. WILLIAMS, *supra* note 58, at 65-67.

61. Green, *supra* note 57, at 653.

62. Power is a function of the study's sample size, the size of the effect one wishes to detect, and the significance criterion used to guard against Type I error. JACOB COHEN, *STATISTICAL POWER ANALYSIS FOR THE BEHAVIORAL SCIENCES* 14 (2d ed. 1988).

63. See generally FREDERIC MARC WOLF, *META-ANALYSIS: QUANTITATIVE METHODS FOR RESEARCH SYNTHESIS* (1986) (basic text on meta-analysis); Thomas R. Einarson et al., *A Method for Meta-Analysis of Epidemiological Studies*, 22 *DRUG INTELLIGENCE & CLINICAL PHARMACY* 813, 813-23 (1988) (applying meta-analysis to Bendectin studies).

64. Thoughtful investigators assess their findings in light of this threat to validity. For example, Shiono and Klebanoff examined births in Northern California for 58 categories of birth defects. Patricia Shiono & Mark Klebanoff, *Bendectin and Human Congenital Malformations*, 40 *TERATOLOGY* 151, 152-55 (1989). Bendectin ingestion was significantly related to three types of defects: lung defects, microcephaly (small head size) and cataracts. *Id.* at 152. The authors noted that three significant relationships out of 58 are "exactly the number of significant relationships that would have been expected by chance" when using 95% confidence intervals, and concluded that the three associations "are unlikely to be causal." *Id.* at 155.

fect. Coding errors occur when researchers treat individuals who were not exposed as having been exposed and those exposed as not exposed, or the researcher misdiagnoses the individuals. Unreliable coding threatens validity by inflating error variance and attenuating true relationships.⁶⁵

B. INTERNAL VALIDITY

Statistical conclusion validity presents a special case of internal validity, which Cook and Campbell define as "the approximate validation with which we infer that a relationship between two variables is causal or that the absence of a relationship implies the absence of cause."⁶⁶ Threats to internal validity usually can be thought of as specification errors. Specification errors occur when the researcher fails to consider a factor that mediates the observed effect between two variables, either because it explains changes in both the "cause" and the "effect" or intervenes between the "cause" and the "effect" and acts independently on the "effect."⁶⁷ Among the threats to internal validity Cook and Campbell discuss are history (the threat that an observed effect may be due to an event that takes place between two points of measurement when this event is not the treatment under investigation),⁶⁸ testing (the threat that an effect may be due to the number of times responses are measured),⁶⁹ and selection (a threat that groups being compared are composed of different types of individuals and, therefore, that observed differences are due to factors other than the treatment under investigation).⁷⁰ A basic advantage of experimental research is its

65. Plaintiff experts in Bendectin cases typically recode data in an attempt to increase reliability. See *infra* part IV.B. (discussing reanalyses of epidemiological data, in particular, those of plaintiff experts Dr. Swan and Dr. Done).

66. COOK & CAMPBELL, *supra* note 56, at 37.

67. See *id.* at 50.

68. *Id.* at 51. Bendectin defendants often introduce evidence indicating that a measurable decrease in birth defects did not accompany rapid withdrawal of the drug from the market between 1981 and 1983. See *infra* note 117 and accompanying text. Discovery of a new teratogenic substance, introduced into the environment over this same period, would constitute a history threat to the validity of the conclusion that Bendectin does not cause birth defects.

69. COOK & CAMPBELL, *supra* note 56, at 52. For example, a prior opportunity to identify a suspect in a lineup may affect the courtroom identification of the suspect.

70. *Id.* at 53. Selection effects take many different forms and often interact with other threats to internal validity. One example of a potential selection effect in Bendectin research derives from the fact that morning sickness is a weak indicator that the fetus is healthy. See Margaret Weigel & Ronald Weigel, *Nausea & Vomiting in Pregnancy and Pregnancy Outcome*, 96 BRITISH J.

ability to control for many selection effects by randomly assigning individuals to treatments.⁷¹ Of course in many situations, such as investigating the effect of toxic substances on individuals, human experiments are impossible. In such cases, the researcher can attempt to control selection threats by carefully matching cases and controls in case-control studies.⁷²

C. CONSTRUCT VALIDITY

The third broad type of validity is construct validity. Confounding operations intended to represent one particular cause or effect construct with some other construct usually threaten construct validity.⁷³ What one investigator may interpret as evidence of a causal relationship between constructs A and B, another investigator may interpret as a relationship between constructs X and B or even X and Y.⁷⁴ There are several sources of construct invalidity. One, experimenter expectancy, occurs when the experimenter anticipates a certain outcome.⁷⁵ Another, evaluation apprehension, arises when the subject wishes to please the investigator.⁷⁶ Finally, hypothesis-guessing may threaten construct validity when the subject attempts to guess the hypothesis being tested and adjust his or her answers accordingly.⁷⁷ For example, drug testing experiments often present construct validity concerns because any observed effect between the drug and a therapeutic effect may be due, not to the

OBSTETRICS & GYNECOLOGY 1312 (1989). Thus, women who took Bendectin were more likely to have a healthy baby than women who did not. *But see* Anne Krickler et al., *Congenital Limb Deficiencies: Maternal Factors in Pregnancy*, 26 AUSTRALIA-NEW ZEALAND J. OBSTETRICS & GYNECOLOGY 272 (1986) (study concluding that "vomiting of (sic) pregnancy was associated with an increased risk of longitudinal limb reduction defects").

71. COOK & CAMPBELL, *supra* note 56, at 55.

72. Even experimental designs cannot control for all threats to internal validity. For example, experiments cannot entirely control for differential mortality in treatment groups. Differential mortality obscures the interpretation of other results because the remaining individuals in the two groups may no longer be comparable on average. COOK & CAMPBELL, *supra* note 56, at 57. This difference may be attributable to the treatment itself, such as when animals die from very large doses of a substance. *See, e.g.*, Rochelle W. Tyl et al., *Developmental Toxicity Evaluation of Bendectin in CD Rats*, 37 TERATOLOGY 539, 540 (1988) (noting high "maternal mortality" in certain rat groups given Bendectin).

73. COOK & CAMPBELL, *supra* note 56, at 59.

74. *Id.*

75. *Id.* at 67.

76. *Id.*

77. *Id.* at 66.

chemical action of the drug, but rather to the psychological expectation that the pill will have a beneficial effect.⁷⁸

Another source of construct invalidity is the confounding of constructs and levels of constructs.⁷⁹ One might conclude that A does not cause B when the test involves very low levels of A. At higher levels of A, however, the researcher might uncover a relationship.⁸⁰ In an attempt to avoid this threat, laboratory animal studies routinely expose animals to suspect drugs at more than one dose level.⁸¹

A similar threat arises whenever one has but a single operationalization of the cause or the effect: a mono-operation bias.⁸² Early animal studies failed to detect the teratogenic effects of Thalidomide, in part because they used species unaffected by the drug.⁸³ Even when there are multiple operationalizations, the use of a single method to measure a relationship threatens validity.⁸⁴ Wherever possible, researchers should employ multiple methods.

Assessing construct validity is frequently a question of convergence and divergence across measures. One is much more likely to believe that a cause and effect relationship exists when different measurements and methods *converge* to produce the same result.⁸⁵ Similarly, researchers are more likely to believe that a cause and effect relationship of a particular type exists if there is a *divergence* between measures and manipulations of related but distinct constructs.⁸⁶

D. EXTERNAL VALIDITY

Finally, there is external validity. Just as statistical conclusion validity is a special type of internal validity, construct va-

78. *Id.* at 61. In an effort to increase construct validity, scientists have designed methods such as placebo controls and double blind designs. *Id.* In a double blind design, neither the subject nor the researcher knows who is receiving the treatment and who is receiving the placebo. *E.g.*, A.G. Hendrickx et al., *Evaluation of Bendectin Embryotoxicity in Non-human Primates: Double-Blind Study in Cynomolgus Monkeys*, 32 *TERATOLOGY* 191 (1985).

79. COOK & CAMPBELL, *supra* note 56, at 67.

80. *Id.*

81. Sanders, *The Bendectin Litigation*, *supra* note 24, at 323.

82. COOK & CAMPBELL, *supra* note 56, at 65.

83. Thalidomide is not a teratogen in all animal species. *See* Max Sherman & Steven Strauss, *Thalidomide: A Twenty-Five Year Perspective*, 41 *FOOD DRUG COSM. L.J.* 458, 461 (1986) (noting that Thalidomide is not a teratogen in rats, mice or hamsters).

84. COOK & CAMPBELL, *supra* note 56, at 66.

85. This is known as convergent validity. *Id.* at 61.

86. This is sometimes called discriminant validity. *Id.*

lidity is a type of external validity. External validity involves the ability to generalize conclusions to *particular* persons, settings and times and to *types* of persons, settings and times.⁸⁷ Cook and Campbell list three basic threats to external validity,⁸⁸ each of which can be expressed in terms of an interaction between a treatment and some other factor. First, the potential interaction between selection and treatment poses a threat to external validity.⁸⁹ If a study uncovers a cause and effect relationship, the researcher must determine to which categories of individuals the relationship can be generalized. For example, if a study includes only men as subjects, the researcher must determine whether the results can be generalized to women. Other examples involve the ability to generalize across race, ethnicity, and class. Other, more subtle threats to generalization may pose special problems in the courtroom. For instance, jurors may rely upon the persuasiveness of experts as an indicator of the merits of their position. Litigants select testifying experts in part for their persuasiveness and, therefore, the assumption that a causal relationship exists between persuasiveness and correctness may be unwarranted for the *type* of expert that appears as a witness in court.⁹⁰

The interaction between setting and treatment creates a second threat. A researcher may not be able to generalize studies done in one setting to other settings. All laboratory studies are vulnerable to this threat.⁹¹ Even well crafted experiments that do their best to increase external validity cannot insure that their results can be transferred from the laboratory. Some laboratory studies suffer from multiple threats to external validity. For example, some laboratory studies of jury decision making involve college sophomores rather than actual jurors.⁹² The subjects read a written fact pattern and each individual "juror" renders his or her own decision, instead of the "jury" issuing a collective decision after deliberation.⁹³ Likewise, laboratory animal studies encounter difficulty in extrapolating across both dose rates and species.⁹⁴ Courts have frequently focused on

87. *Id.* at 71.

88. *Id.* at 73-74.

89. *Id.* at 73.

90. Samuel R. Gross, *Expert Evidence*, 1991 WIS. L. REV. 1113, 1134 (1991).

91. COOK & CAMPBELL, *supra* note 56, at 74.

92. REID HASTIE ET AL., *INSIDE THE JURY* 40 (1983).

93. *Id.*

94. *See infra* notes 125-131 and accompanying text.

threats to external validity when refusing to admit expert testimony.⁹⁵

This brief review of validity's different facets indicates some of the ways in which conclusions about causation may be in error. Statistical conclusion validity and other types of internal validity concentrate on the danger of Type I or Type II errors, drawing false positive or false negative conclusions about causation. Statistical conclusion validity deals with threats to internal validity caused by random error, the possibility that an observed relationship could be due to chance. Other threats to internal validity are due to the possible existence of bias through factors that systematically affect the value of the means of variables.⁹⁶ Construct validity and other types of external validity concentrate on the danger of generalization. The principal threat stems from the possible existence of an undetected interaction.⁹⁷ With respect to construct validity the danger is that an effect can be obtained using one measure, such as individual juror judgments, and a different effect using a different measure, such as collective jury judgments. The risk of undetected interaction effects is even easier to see with respect to other threats to external validity, such as the interaction between selection and treatment.⁹⁸ In each case, a relationship observed in one circumstance may not apply in a different circumstance. The next section employs these categories of scientific validity to discuss specific admissibility rulings in Bendectin cases.

IV. RESTRICTIVE ADMISSIBILITY RULINGS IN THE BENDECTIN CASES

Bendectin has become a very important product, primarily because it has precipitated a reanalysis of the judiciary's proper role in assessing the admissibility of scientific evidence. For nearly a decade, trial and appellate courts have wrestled with the admissibility of plaintiffs' expert testimony that Bendectin is

95. For example, several courts, like the *Daubert* trial court, have resisted the introduction of expert opinion based on animal studies because of concerns about external validity. See *infra* text accompanying notes 125-29 (discussing the external validity problem posed by animal studies); *infra* text accompanying notes 134-48 (discussing courts' refusal to allow non-epidemiological evidence such as animal studies).

96. COOK & CAMPBELL, *supra* note 56, at 80.

97. *Id.* at 81.

98. *Id.* at 74.

a teratogen.⁹⁹ This section investigates whether *Daubert's* scientific validity analysis can explain and justify the Bendectin cases' restrictive admissibility rulings and whether the excluded testimony in those cases was so invalid that the courts properly excluded it. The Article addresses this question in the context of two specific themes that have arisen in Bendectin litigation: the primacy of epidemiology, and the exclusion of testimony based on a reanalysis of published epidemiological results.

A. THE PRIMACY OF EPIDEMIOLOGY

1. Evidence of a Causal Relationship

Bendectin is a substance that is not obviously harmful.¹⁰⁰ It does not produce a signature disease and no generally accepted biological theory exist about how it produces its alleged effect. Moreover, the correlation between the product and the plaintiffs' injuries is not strong. The evidence as to the causal relationship between the drug and birth defects comes in five basic types: structure-activity, *in vitro* research, animal studies, epidemiology, and secular trend analysis.¹⁰¹

Structure-activity. Substances with similar chemical structures may have similar effects on the human body. Bendectin contains doxylamine succinate, an antihistamine acting as an antinauseant.¹⁰² Some antihistamines are known teratogens and plaintiff experts point to this structural similarity.¹⁰³

In vitro. *In vitro* research involves exposing cells or organs maintained in a culture to a substance.¹⁰⁴ One study indicates that Bendectin inhibits certain limb bud cell differentiation¹⁰⁵

99. A list of all reported Bendectin opinions through 1991 can be found in Sanders, *The Bendectin Litigation*, *supra* note 24, at 410-18.

100. See Louis Lasagna & Sheila R. Shulman, *Bendectin and the Language of Causation*, in PHANTOM RISK: SCIENTIFIC INFERENCE AND THE LAW 100, 100-01 (Kenneth R. Foster et al. eds., 1993). Other mass exposures that currently fit into this category include exposure to Agent Orange, PCBs, certain toxic waste dumps, breast implants, and electro-magnetic fields.

101. Joseph Sanders, *The Jury Deliberations in a Complex Case: Havner v. Merrell Dow Pharmaceuticals, Inc.*, 16 JUST. SYS. J. 45, 52-53 (1993) [hereinafter Sanders, *Jury Deliberation*].

102. Sanders, *The Bendectin Litigation*, *supra* note 24, at 317.

103. DeLuca v. Merrell Dow Pharmaceuticals, Inc., 911 F.2d 941, 946 n.8 (3d Cir. 1990).

104. See, e.g., Stuart Freeman et al., *Post-implantation Embryo Culture for Studies of Teratogenesis*, in BIOCHEMICAL TOXICOLOGY: A PRACTICAL APPROACH 83 (K. Snell & B. Mullock eds. 1987) (assessing studies of teratogenesis).

105. John R. Hassell & Elizabeth A. Horigan, *Chondrogenesis: A Model Developmental System for Measuring Teratogenic Potential of Compounds*, 2 TERATOGENESIS, CARCINOGENESIS, & MUTAGENESIS 325, 325-27 (1982).

and another suggests that Bendectin may be a weak DNA damaging agent.¹⁰⁶

Animal studies. *In vivo* studies examine the effects of a substance on animals. Researchers have conducted Bendectin animal studies on chicks, rats, rabbits, and primates.¹⁰⁷ As plaintiff experts note, some studies report a relationship between Bendectin or one of its ingredients and a teratogenic effect. For example, one primate study found that the drug caused a delay in the closure of the ventricular septa.¹⁰⁸ On the other hand, several studies have failed to find a correlation between Bendectin and birth defects.¹⁰⁹

Epidemiology. Epidemiological studies compare the incidence of birth defects among those exposed to and those not exposed to a substance. There are two general ways of making such comparisons: cohort studies and case-control studies.¹¹⁰ Cohort studies compare the incidence of defects among groups of persons exposed to the substance and groups of persons not exposed.¹¹¹ Case-control studies match a group of persons who have the injury in question with another group that does not have that injury.¹¹² The studies then compare exposure rates for the two groups. Nearly 40 published epidemiological studies discuss Bendectin.¹¹³ In no individual study did the authors conclude that Bendectin is a teratogen. In six studies, however, the authors found at least one significant correlation between Bendectin use and some injury and concluded that, although a single study alone is insufficient to support an attribution of causation, an effect might exist.¹¹⁴ In the remaining 33 studies

106. John D. Budroe et al., *A Study of the Potential Genotoxicity of Methapyrilene and Related Antihistamines using the Hepatocyte/DNA Repair Assay*, 135 *MUTATION RES.* 131, 135-36 (1984).

107. Sanders, *The Bendectin Litigation*, *supra* note 24, at 394.

108. A.G. Hendrickx et al., *Evaluation of Bendectin Embryotoxicity in Non-human Primates: Ventricular Septal Defects in Prenatal Macaques and Baboon*, 32 *TERATOLOGY* 179, 179-89 (1985); A.G. Hendrickx et al., *supra* note 78, at 194.

109. For a list of the published animal studies, see Sanders, *The Bendectin Litigation*, *supra* note 24, at 403.

110. Leon Gordis, *Estimating Risk and Inferring Causality in Epidemiology*, in *EPIDEMIOLOGY AND HEALTH RISK ASSESSMENT* 51, 52 (Leon Gordis ed. 1988).

111. *Id.*

112. *Id.*

113. For a list of published epidemiological studies through 1991, see Sanders, *The Bendectin Litigation*, *supra* note 24, at 404-06.

114. Pamela Aselton et al., *Pyloric Stenosis and Maternal Bendectin Exposure*, 120 *AM. J. EPIDEMIOLOGY* 251 (1984); Jose F. Cordero et al., *Is Bendectin a Teratogen?*, 245 *JAMA* 2307 (1981); Brenda Eskenazi & Michael B. Bracken, *Bendectin (Debendox) as a Risk Factor for Pyloric Stenosis*, 144 *AM. J. OBSTET-*

the authors either drew no conclusion or concluded that no statistical relationship existed.¹¹⁵

Secular trend. Secular trend data, which is similar to epidemiology, compares the total reported incidence of various types of birth defects with the volume of Bendectin sales and prescriptions. This method investigates whether increases or reductions in birth defects paralleled the rapid increase in Bendectin prescriptions in the 1970s or the precipitous drop in prescriptions in the early 1980s.

In recent years defense experts have concentrated their testimony on the epidemiological and secular trend evidence. As to the epidemiological data, Bendectin defendants have argued that, taken as a group, the studies indicate that Bendectin is not teratogenic.¹¹⁶ They have also reviewed secular trend evidence that indicates no significant decrease in birth defects after Bendectin manufacturers withdrew the product from the mar-

RICS & GYNECOLOGY 919 (1982); G.T. Gibson et al., *Congenital Anomalies in Relation to the Use of Doxylamine/Dicyclomine and other Antenatal Factors: An Ongoing Prospective Study*, 1981 MED. J. AUSTRALIA 410 (1981); J. Golding et al., *Maternal Anti-nauseants and Clefts of Lip and Palate*, 1983 HUMAN TOXICOLOGY 63 (1983); Kenneth J. Rothman et al., *Exogenous Hormones and Other Drug Exposures of Children with Congenital Heart Disease*, 109 AM. J. EPIDEMIOLOGY 433, 435 (1979).

115. Sanders, *The Bendectin Litigation*, *supra* note 24, at 395.

116. Einarson, Leeder & Koren included 17 studies in a meta-analysis examining whether first-trimester Bendectin ingestion caused any birth defect. See Einarson et al., *supra* note 63. The overall odds ratio was 1.01, $\chi^2 = 0.05$, $p = 0.815$. They also conducted separate meta-analyses for cohort and case control studies. For cohort studies ($N = 12$) the ratio was 0.95, $\chi^2 = 0.66$, $p = 0.418$. For case control ($N = 5$) studies the ratio was 1.27, $\chi^2 = 2.71$, $p = 0.10$. *Id.* at 819-20. The authors concluded that these meta-analyses confirm previous subjective analyses that Bendectin is not associated with human teratogenic outcomes. *Id.* at 822; see also Leslie J. Sheffield & Ron Batagol, *The Creation of Therapeutic Orphans—Or, What Have We Learnt From the Debendox Fiasco?*, 143 MED. J. AUSTRALIA 143, 144-45 (1985) (noting “great uniformity” in studies “finding no teratogenic effect of Debendox”).

The odds ratio is the cross product in a 2 x 2 table. In a cohort study examining exposed and unexposed individuals, the odds ratio is the ratio of the odds of injury if the person was exposed, to the odds of injury if the person was not exposed.

ODDS RATIO IN A COHORT STUDY

		Not Injured		
	Injured	b	c	
Exposed	a	b	c	$\frac{a/b}{c/d} = \frac{ad}{bc}$
Not Exposed	c	d	d	

In a case-control study comparing injured subjects to “controls” without the injury, the odds ratio is the ratio of the odds that the injured subjects suffered exposure to the odds that the controls suffered exposure.

ket.¹¹⁷ Plaintiff experts, on the other hand, have devoted substantial attention to structure-activity, *in vitro*, and animal studies, as well as to reanalyses of epidemiological data.¹¹⁸

2. Problems with Non-Epidemiological Evidence

Structure-activity, *in vitro*, and animal studies each pose substantial validity questions. Most problematic, perhaps, is the structure-activity evidence. Although research exists linking antihistamines to teratogenic injuries, several factors undermine its validity.¹¹⁹ First, even minor changes in molecular structure can alter a substance's effect.¹²⁰ The metabolic process stands as an unknown intervening variable between the original chemical structure and the adverse effect. Thus, structure-activity data presents a problem of internal validity.

In vitro evidence is superior to structure-activity evidence because it does investigate the effect of the Bendectin ingredients. *In vitro* evidence suffers, however, from the same internal validity problem confronting structure-activity data because the relevant chemical compound does not go through the metabolic process before affecting the culture.¹²¹ Moreover, the confounding of constructs and levels of constructs threatens this type of evidence. For example, the study which found that Bendectin inhibited cell differentiation employed a unit of measure called the teratogenic potential.¹²² The authors observed

ODDS RATIO IN A CASE-CONTROL STUDY

	Cases (With Injury)	Controls (Without Injury)	
History of Exposure	a	b	$\frac{a}{c} = \frac{ad}{bc}$
No History of Exposure	c	d	

For example, if one conducted a case-control study with the following results $a=60$, $b=40$, $c=40$, $d=60$, the odds ratio would be $(60 \cdot 60) / (40 \cdot 40) = 3600 / 1600 = 2.75$. See HAROLD KAHN, AN INTRODUCTION TO EPIDEMIOLOGIC METHODS 38-45 (1983).

117. Robert L. Brent, *Bendectin and Interventricular Septal Defects*, 32 TERATOLOGY 317, 317 (1985); see D.W.G. Harron et al., *Debendox and Congenital Malformations in Northern Ireland*, 281 BRIT. MED. J. 1379, 1381 (1980).

118. Sanders, *From Science to Evidence*, *supra* note 25, at 43-44.

119. C.T.G. King et al., *Antihistamines and Teratogenicity in the Rat*, 147 J. PHARMACOLOGY & EXPERIMENTAL THERAPEUTICS 391, 395 (1965).

120. Green, *supra* note 57, at 658.

121. O.P. Flint, *An In Vitro Test for Teratogens Using Cultures of Rat Embryo Cells*, reprinted in IN VITRO METHODS IN TOXICOLOGY 339, 354 (C.K. Atterwill & C.E. Steele eds. 1987).

122. Hassell & Horigan, *supra* note 105, at 327, 330.

an effect for Bendectin at a dose of .05 mg/ml.¹²³ Caffeine produces a similar effect at 2.3mg/ml and vitamin A does so at .000013 mg/ml.¹²⁴ It is difficult to translate such dosages in *in vitro* studies to the doses humans actually experience.

Because animal studies require ingestion of a drug, they do not confront all of the threats to internal validity that structure-activity and *in vitro* studies face. They do, however, confront external validity threats. Some of the threats result inevitably from reasonable tradeoffs designed to avoid threats to internal validity. Among these tradeoffs, dose rates pose the most important problem. Researchers usually give animals a substance at a dose rate much higher than humans would ingest.¹²⁵ Several compelling reasons merit this practice. Animal research is expensive and time consuming. Many substances that are suspected of causing harm do so in only in a small percentage of organisms exposed at a rate similar to that found in the environment.¹²⁶ Subjecting the animals to dose rates no greater than typical environmental rates would require a very large N to avoid a high probability of a Type II error. Thus, in order to guard against threats to statistical conclusion validity, researchers increase the dose so that a larger percentage of animals will react adversely.¹²⁷ This closage, however, creates a significant threat to external validity. The high doses create the potential for construct validity problems similar to those presented in *in vitro* tests: confounding constructs with levels of constructs. At sufficiently high dose levels almost all substances are teratogenic.¹²⁸ Moreover, in the case of suspected teratogens, very high animal dose rates begin to poison the mother and cause fetal injuries as the byproduct of maternal toxicity and not the substance's teratogenic effect.¹²⁹

123. *Id.* at 330.

124. *Id.* The authors do not express an opinion about whether Bendectin is dangerous to humans when taken in normal therapeutic doses. *Id.* at 330-31.

125. Michael D. Hogan & David G. Hoel, *Extrapolation to Man*, reprinted in *PRINCIPLES & METHODS OF TOXICOLOGY* 879, 879-80 (Wallace Hayes ed., 2d ed. 1989).

126. Jack L. Landan & 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, 537 (1988-89).

127. Even with high dose rates, the relatively small number of animals in some experiments may create a threat to statistical conclusion validity when searching for a weak causal link.

128. Thomas H. Shepard, *Human Teratogenicity*, 33 *ADVANCES IN PEDIATRICS* 225, 227 (1986).

129. See Tyl et al., *supra* note 72, at 549.

Even at more modest dose rates, extrapolation difficulties pose significant threats to external validity. Assuming a positive result in an animal study, toxicologists must then extrapolate a predicted effect at a dose level humans actually experience. No single agreed upon model for this extrapolation exists and competing models produce different predictions.¹³⁰ Nor is dose rate the only necessary adjustment. There must also be an adjustment for the fact that species are of different sizes and mature and age at different rates. Again, there is no agreed upon formula for this adjustment, and different scaling factors lead to different estimates of human effects.¹³¹

One reason toxicologists tolerate the threat to external validity posed by high dose rates is that most animal studies are designed to be part of the regulatory process rather than part of proof of causation in litigation. When testing a new drug the critical question is whether a teratogenic effect might arise in humans even though it is not observed in animals. The crucial error to avoid is a Type II error. When litigants take these studies to the courtroom, however, the central question becomes whether a known effect in a test animal is probative of whether a human effect exists at a much lower dose rate. Although it is quite rare for a known human teratogen to fail to cause birth defects in at least some animals,¹³² it is more likely that a substance for which there is no evidence of human teratogenicity will produce an effect in some animal species.¹³³

130. For example, varying statistical models for extrapolating carcinogenic effects produce different results when the laboratory dose rate is substantially greater than the environmental dose rate. OSHA Generic Cancer Policy, 45 Fed. Reg. 5002, 5184-85 (1980); David S. Salsburg, *Statistics and Toxicology: An Overview*, in SCIENTIFIC CONSIDERATIONS IN MONITORING AND EVALUATING TOXICOLOGICAL RESEARCH 123, 130-31 (Edward Gralla ed., 1981).

131. See James P. Leape, *Quantitative Risk Assessment in Regulation of Environmental Carcinogens*, 4 HARV. ENVTL. L. REV. 86, 98-99 (1980). Comparisons of risk estimates of cancer based on extrapolations from animal data with actual human epidemiological data indicate that only half the substances examined yielded accurate estimates. Landau & O'Riordan, *supra* note 126, at 548 (indicating that many estimates err by factors of 10 or more).

132. IAN C.T. NISBET & NATHAN J. KARCH, CHEMICAL HAZARDS TO HUMAN REPRODUCTION 98-99 (1983).

133. A 1980 FDA study reported that of 165 compounds with no reported human teratologic effects, only 28% appeared negative in all animal species tested. 45 Fed. Reg. 69,816, 69,823 (1980); NISBET & KARCH, *supra* note 132, at 105. For explanations of how and why effects in humans differ from various animal species, see EDWARD J. CALABRESE, PRINCIPLES OF ANIMAL EXTRAPOLATION 237-38 (1983); Gary P. Carlson, *Factors Modifying Toxicity*, in TOXIC SUBSTANCES AND HUMAN RISK: PRINCIPLES OF DATA INTERPRETATION 47, 49 (R. Tardiff & J. Rodricks eds., 1987). One example is reported in *Turpin v. Merrell*

3. Court Attitudes Towards Non-Epidemiological Data

The validity threats facing non-epidemiological data have caused several courts to refuse to allow Bendectin plaintiffs to introduce this type of evidence or prevail thereon.¹³⁴ Courts have reached this conclusion by a number of different paths. One group has directed verdicts or ordered summary judgment for the defendant without ruling on the admissibility of the plaintiff's non-epidemiology experts.¹³⁵ These are properly characterized as sufficiency rulings; the court concludes that the plaintiff cannot survive a directed verdict or a summary judgment motion because the causal proof cannot sustain a verdict for the plaintiff. For example, in *Brock v. Merrell Dow*, the Fifth Circuit held that the plaintiff could not prevail without epidemiological evidence of a statistically significant relationship between Bendectin and the plaintiff's limb reduction defect.¹³⁶ Because the opinion followed a jury trial, the Fifth Circuit did not need to rule on the admissibility of the plaintiff's non-epidemiological evidence.¹³⁷ In *Turpin v. Merrell Dow Pharmaceuticals, Inc.*,¹³⁸ however, the Sixth Circuit affirmed the trial court's grant of summary judgment for the defendant and held that the plaintiff's proffered testimony was insufficient to sustain a verdict.¹³⁹

Dow Pharmaceuticals, Inc., 959 F.2d 1349, 1359 n.4 (6th Cir. 1992), *cert. denied*, 113 S. Ct. 84 (1992). Several animal studies have found that cortisone causes severe cleft palate birth defects in several animal species, but not in humans. Alfred M. Bongiovanni & Arthur J. McPadden, *Steroids During Pregnancy and Possible Fetal Consequences*, 11 FERTILITY & STERILITY 181, 184-85 (1960). It may be, of course, that some chemicals with no apparent carcinogenic or teratogenic effect are not in fact completely harmless because of significant limits on the ability of epidemiological studies to detect small risks.

134. See *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 727 F. Supp. 570, 575 (S.D. Cal. 1989), *aff'd*, 951 F.2d 1128 (9th Cir. 1991), *vacated*, 113 S. Ct. 2786 (1993).

135. See, e.g., *Brock v. Merrell Dow Pharmaceuticals, Inc.*, 874 F.2d 307, 311-13 (5th Cir. 1989), *cert. denied*, 494 U.S. 1046 (1990).

136. *Id.* at 313.

137. *Id.* at 313-15; see *Merrell Dow Pharmaceuticals, Inc. v. Havner*, No. 13-92-540-CV, 1994 WL 86436 (Tex. Ct. App. March 17, 1994)

138. 959 F.2d 1349 (6th Cir.), *cert. denied*, 113 S. Ct. 84 (1992).

139. *Id.* at 1360-61. The court refused to conclude that animal studies could never form the basis of an opinion that a substance is a human teratogen, only that the plaintiff's animal studies could not. *Id.* at 1360. The *Turpin* district court excluded much of the plaintiff's evidence under Rule 703 but held in the alternative that, even if admissible as a matter of law, the evidence could not support a verdict for the plaintiff. *Turpin v. Merrell Dow Pharmaceuticals, Inc.*, 736 F. Supp. 737, 744 (E.D. Ky. 1990); see also *Elkins v. Richardson-Merrell, Inc.*, 8 F.3d 1068 (6th Cir. 1993) (affirming grant of summary judgement for defendant by relying on *Turpin*), *cert. denied*, 62 USLW 3618 (1994).

Another group of courts simply holds that non-epidemiological evidence is inadmissible.¹⁴⁰ Some, such as the Ninth Circuit in *Daubert*, reach this result under the *Frye* test.¹⁴¹ More relevant in the post-*Frye* environment, however, are those cases which excluded the plaintiff's testimony under Federal Rule 702 or 703.¹⁴² In *Lynch v. Merrell-National Laboratories*,¹⁴³ one of the earliest opinions to take this position, the trial court entered summary judgment for the defendant after concluding that testimony on human teratogenicity based on structure-activity, *in vitro*, or animal studies was not of the "type reasonably relied upon by experts in the particular field" and therefore inadmissible.¹⁴⁴ A similar analysis can be found in *Richardson v. Richardson-Merrell, Inc.*:

140. See, e.g., *Richardson v. Richardson-Merrell, Inc.*, 857 F.2d 823 (D.C. Cir. 1988), cert. denied, 493 U.S. 882 (1989).

141. *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 951 F.2d 1128 (9th Cir. 1991), vacated, 113 S. Ct. 2786 (1993).

142. An important non-Bendectin case with a similar ruling is *In re "Agent Orange" Products Liability Litigation*. 611 F. Supp. 1223 (E.D.N.Y. 1985), *aff'd*, 818 F.2d 145 (2d Cir. 1987), cert. denied, 487 U.S. 1234 (1988). There, Judge Weinstein refused to allow the plaintiff's experts to base their opinion on animal studies, primarily because of external validity concerns:

The many studies on animal exposure to Agent Orange, even plaintiffs' expert concedes, are not persuasive in this lawsuit There is no evidence that plaintiffs were exposed to the far higher concentrations involved in both the animal and industrial exposure studies. Cf. *In re "Agent Orange" Product Liability Litigation*, 597 F. Supp. 740, 782 (E.D.N.Y.1984). The animal studies are not helpful in the instant case because they involve different biological species. They are of so little probative force and are so potentially misleading as to be inadmissible. See FED. R. EVID. 401-403. They cannot be an acceptable predicate for an opinion under Rule 703.

In re "Agent Orange", 611 F.Supp. at 1241; see *In re Paoli Railroad Yard PCB Litigation*, Nos. 86-2229, 1992 U.S. Dist. LEXIS 16287, at *15 (E.D. Pa. Oct. 21, 1992) (holding that animal studies are of limited reliability when attempting to apply their results to humans).

143. 646 F. Supp. 856 (D. Mass. 1986), *aff'd*, 830 F.2d 1190 (1st Cir. 1987).

144. *Id.* at 866-67. The court noted the following:

None of the animal studies submitted by the plaintiffs provide evidence of teratogenicity at doses comparable to the human therapeutic dose of Bendectin. These animal studies are therefore lacking in probative value and must be found inadmissible. . . . For similar reasons, this Court must reject the plaintiffs' proffered evidence of *in vitro* studies and studies of analogous chemical structures as a basis for the plaintiffs' experts testimony. . . . This Court also cannot find, pursuant to Rule 703, that such studies are "of a type reasonably relied upon by experts in the particular field." Dr. John Hassell, the author of one such *in vitro* study, has expressly recognized that neither his technique nor any other *in vitro* system has yet been validated as an accurate predictor of teratogenicity in animals or humans. . . . Thus, a careful review of the material before this Court indicates that the only relevant, probative, and non-misleading evidence on the issue of Bendec-

These three types of studies then—chemical, *in vitro*, and *in vivo*—cannot furnish a sufficient foundation for a conclusion that Bendectin caused the birth defects at issue in this case. Studies of this kind, singly or in combination, are not capable of proving causation in human beings in the face of the overwhelming body of contradictory epidemiological evidence.¹⁴⁵

The key to *Richardson* is its comparative analysis. The court essentially held that structure-activity, *in vitro* and animal studies cannot form a sufficient foundation when substantial epidemiological evidence exists.¹⁴⁶ Similar language can be found in the district court opinions in *Lee v. Richardson-Merrell, Inc.*,¹⁴⁷ and *Turpin v. Merrell Dow Pharmaceuticals, Inc.*,¹⁴⁸ both of which excluded non-epidemiological evidence.

Whether such non-epidemiological evidence should be excluded under a scientific validity standard turns on how precisely one poses the issue. If one asks whether, standing alone, structure-activity, *in vitro* and animal studies should be excluded, the answer depends on which type of evidence is under consideration. Inevitably, validity is a matter of degree. All types of non-epidemiological evidence suffer from some external validity problems when used to address whether Bendectin is a human teratogen at normal dose levels. The animal studies data, however, confronts far fewer problems and have some strengths vis-a-vis epidemiological studies. If the only evidence

tin's role in the causation of birth defects are the controlled observations of human beings, documented in more than 25 published epidemiological studies.

Id.

145. 857 F.2d 823 (D.C. Cir. 1988).

146. In situations where there is not a substantial body of epidemiological data, however, courts have been more accepting of animal study data. See, e.g., *Marder v. G.D. Searle & Co.*, 630 F. Supp. 1087, 1094 (D. Md. 1986), *aff'd sub nom.*, *Wheelahan v. G.D. Searle & Co.*, 814 F.2d 655 (4th Cir. 1987).

147. 772 F. Supp. 1027, 1029-33 (W.D. Tenn. 1991), *aff'd*, 961 F.2d 1577 (6th Cir.), *cert. denied*, 113 S. Ct. 197 (1992).

148. 736 F. Supp. 737, 739-44 (E.D. Ky. 1990). The courts in *Lee* and *Turpin* employed a test developed in *United States v. Green*, 548 F.2d 1261 (6th Cir. 1977) and *United States v. Kozminski*, 821 F.2d 1186 (6th Cir. 1987), *aff'd*, 487 U.S. 931 (1988). The *Green* court set forth a four-prong test for admissibility: (1) a qualified expert must be offered; (2) the expert must testify on a proper subject; (3) the expert must testify in conformity with a generally accepted explanatory theory; and (4) the probative value of the testimony must outweigh any prejudicial effect. *Green*, 548 F.2d at 1268. *Kozminski* further refined the third element by requiring that the explanatory theory must have: (a) received at least some exposure within the scientific peerage to which it belongs; (b) been subjected to peer evaluation to determine its scientific validity and reliability; and (c) achieved general acceptance within the scientific community to which it belongs. *Kozminski*, 821 F.2d at 1201.

available to a "first plaintiff"¹⁴⁹ expert is an animal study indicating a teratogenic effect on mice exposed to Bendectin at a dose rate one order of magnitude greater than a human dose, it is difficult to see why this should not be admissible. The question becomes more difficult when the only evidence is an *in vitro* study indicating DNA damage to cells exposed to Bendectin. In an extreme case, where the only available evidence is a structure-activity study relating some antihistamines to birth defects, the threats to internal and external validity may indeed be so large that the evidence cannot form the basis of an expert opinion that Bendectin is a teratogen. Thus, judging each type of evidence on its own, a court might reasonably exclude an expert's conclusion that Bendectin is a teratogen if it is based solely on structure-activity evidence, but might admit an opinion based on animal studies.

The courts that have excluded non-epidemiological data in the Bendectin cases have not approached the problem in this way. They have not independently assessed the admissibility of each type of evidence as if it were the only available evidence and they been hesitant to conclude that plaintiffs can never reach a jury without epidemiological evidence. Instead, they have carved out an exception for Bendectin cases because of the rich epidemiological data available.¹⁵⁰ Whether the admissibil-

149. A "first plaintiff" is the first individual to claim that a toxic substance causes injury. This individual must frequently litigate on an undeveloped scientific record. See Sanders, *The Bendectin Litigation*, *supra* note 24, at 349.

150. See *Richardson v. Richardson-Merrell, Inc.*, 857 F.2d 823 (D.C. Cir. 1988). At least two courts have refused to adopt this unique approach in Bendectin cases. In *Longmore v. Merrell Dow Pharmaceuticals, Inc.*, the court concluded as follows:

Animal studies are generally relied upon by experts determining the link between a drug and birth defects and the same is true for chemical analysis. While the Court will leave open the question of the admissibility of particular studies during the trial of this matter, the Court cannot now preclude all such studies under Rule 703.

737 F. Supp. 1117, 1121 (D. Idaho 1990). In *In Re Bendectin Products Liability Litigation*, Judge Rubin, the trial judge who presided over the 1985 Multidistrict Litigation Bendectin Trial in Ohio, also refused to hold non-epidemiological evidence inadmissible under Rule 703, noting the following:

The division in the scientific community over whether epidemiological studies should be relied upon exclusively necessitates the inescapable conclusion that experts may reasonably rely upon other types of data when forming an opinion as to the teratogenicity of Bendectin. A contrary finding is unjustifiable without a pronouncement in this circuit that, as a matter of law, epidemiological studies are the sole basis upon which an expert may reasonably rely when forming an opinion on a drug's teratogenicity.

732 F. Supp. 744, 749 (E.D. Mich. 1990) (citations omitted).

ity of one type of evidence should vary depending on the existence of other evidence poses an interesting, but different, question. More precisely, should the admissibility of non-epidemiological evidence turn on the existence of epidemiological evidence? Applying a validity analysis under Rule 702, the answer is no. Under that standard, if the threats to internal and external validity do not render a piece of evidence unreliable, that evidence does not become unreliable simply because better data is available.

The Bendectin courts, however, have not relied on Rule 702 as the basis of their opinions. Rather, most have found the evidence inadmissible under Rule 703 and held that non-epidemiological findings are not the type of evidence relied upon by experts in the field.¹⁵¹ The propriety of this approach depends on the structure of the plaintiff's entire case. An everyday example illustrates this point. When the only evidence whether a dog walked across the front lawn last night is the report of three eyewitnesses who say they saw no dog, this constitutes the best evidence available and is admissible. When, on the other hand, it snowed during the night, the presence of dog tracks across the lawn greatly diminishes the value of the eyewitness testimony. Similarly, when animal studies supply the only available evidence that a substance causes harm, teratology experts may reasonably rely on this finding. When a large and rich body of epidemiological data exists, however, experts may cease to rely on the animal study as the primary basis of their opinion about whether the substance is a teratogen. Nevertheless, an independent Rule 702 validity analysis indicates why courts should refuse to conclude that non-epidemiological evidence is inadmissible even when a large body of epidemiological data exists.

One could argue, under Rule 703, that it would be inappropriate to form one's opinion about whether Bendectin is a teratogen based entirely on non-epidemiological evidence. Something like a "best scientific evidence" rule might render an opinion based solely on such evidence inadmissible.¹⁵² This does *not* mean, however, that the best alternative is to base an opinion solely on the epidemiological evidence. On the contrary, from a scientific validity perspective it would be preferable to

151. See, e.g., *Richardson*, 857 F.2d at 829-32.

152. See *Green*, *supra* note 57, at 676 (noting that the probity of toxicology evidence, especially animal studies, varies inversely with the quality of the epidemiological evidence).

form one's opinion about Bendectin based on all of the available evidence: epidemiological, animal studies, *in vitro* and perhaps even structure-activity. This conclusion focuses on questions of construct validity. Any single operation—any single study—is threatened by a potential mono-operation bias.¹⁵³ Cook and Campbell note: "Since single operations both under represent constructs and contain irrelevancies, construct validity will be lower in single exemplar research than in research where each construct is multiple operationalized in order to triangulate on the referent."¹⁵⁴ Thus, confounders or other irrelevancies, may affect the results of a single epidemiological study. Multiple studies provide greater certainty that the observed relationships does in fact represent the concepts under investigation.

Even multiple epidemiological replications suffer from mono-method bias. "[W]hen all the manipulations are presented in the same way, or all the measures use the same means of recording responses, then the method is itself an irrelevancy whose influence cannot be dissociated from the influence of the target construct."¹⁵⁵ As a method, epidemiology has many strengths, but it also has weaknesses. Because it is not an experimental method, it inevitably suffers from some internal validity problems. On the other hand, because animal studies are experiments, it is relatively more certain that the substance experimentally manipulated caused any observed effect. Each type of evidence addresses some of the weaknesses of the other. The whole is greater than the sum of the parts.

Bendectin plaintiffs have never argued that epidemiological data should be disregarded. On the contrary, plaintiff experts who are prepared to testify that more likely than not Bendectin caused the plaintiff's birth defect, have based their conclusion on all available evidence, including animal studies and epidemiological research.¹⁵⁶ Excluding non-epidemiological evidence because better, epidemiological evidence exists is erroneous under a scientific validity standard. Within the context of a scientific validity discussion, the admissibility of such evidence is not contingent upon the existence of other, arguably better evidence. Likewise, exclusion under Rule 703 conflicts with common scientific understandings of construct validity and mono-method bias.

153. COOK & CAMPBELL, *supra* note 56, at 65.

154. *Id.*

155. *Id.* at 66.

156. See, e.g., *Oxendine v. Merrell Dow Pharmaceuticals, Inc.*, 506 A.2d 1100, 1104-08 (D.C. Cir. 1986).

B. THE EXCLUSION OF THE REANALYSIS OF EPIDEMIOLOGICAL STUDIES

Plaintiff efforts to produce a prima facie case of causation have never depended entirely on non-epidemiological evidence. Throughout, Bendectin plaintiffs have offered epidemiological evidence in the form of a reanalysis of existing epidemiological studies.¹⁵⁷ Epidemiological studies have typically escaped the criticism that they are so invalid as to be inadmissible. Nevertheless, plaintiff experts usually design the reanalyses to correct for alleged threats to the validity of results reported in published epidemiological research.¹⁵⁸ They have focused on threats to statistical conclusion validity, primarily stemming from unreliable measurement and the relatively small Ns of many studies.¹⁵⁹ The most serious measurement threat derives from the fact that many of the studies were unable to determine exactly when the expectant mothers took Bendectin. For example, limb reduction defects occur when the limbs are first forming, a period that lasts approximately two weeks.¹⁶⁰ If researchers included women who took Bendectin after this period among the exposed group, the study will underestimate any effect.¹⁶¹ The effect of changes in defining exposure can be dramatic.¹⁶² The existence of doxylamine succinate in products such as Unisom adds another potential source of bias.¹⁶³ Unless researchers ask women whether they took such products during pregnancy, they may code certain women as unexposed who in fact ingested Bendectin's most suspect ingredient. As in the case of misclassification due to time of ingestion, coding errors will underestimate any effect.

157. See Testimony of Dr. Shanna Helen Swan, morning session, Sept. 19, 1991, at 43-64, *Havner v. Merrell Dow Pharmaceuticals, Inc.*, No. 88-3915-F (Tex. Dist. Ct., 214th Jud. Dist., March 17, 1994) (on file with the author).

158. *Id.*

159. *Id.*

160. Green, *supra* note 57, at 650.

161. When it is unknown whether the mother ingested the drug during organogenesis, the study will mistakenly categorize women who took the drug too late in their pregnancy for it to cause a defect as women exposed to Bendectin. These women should be counted as "controls," women not exposed to the drug. The precise consequences of this misclassification are difficult to assess. For "negative cases"—women whose children do not have defects—the misclassification underestimates the drug's teratogenic. For "positive cases," however, the misclassification overestimates the teratogenic effect. Overall, such misclassification introduces an error term that will attenuate any effect that does exist. See Green, *supra* note 57, at 650 n.32.

162. See *id.* at 650.

163. PHYSICIAN'S DESK REFERENCE 2642 (1994).

In addition, because limb reduction defects, which underlie many Bendectin claims, are rare events, the number of individuals who suffer from a given defect and whose mothers took Bendectin is relatively small. As a consequence, there is a substantial risk of making a Type II error due to low statistical power. For example, five cohort studies which report limb reduction injuries together contain only eleven such cases.¹⁶⁴ One way to increase statistical power is to conduct case-control studies that purposefully pick as cases individuals who exhibit the injury under investigation.¹⁶⁵ Four published case-control studies include limb reduction defects with a total N of 312 cases, 61 of whom had mothers who were exposed to the drug.¹⁶⁶ Only two of the studies had Ns sufficiently large to afford a 50% chance of detecting a relative risk of two or smaller.¹⁶⁷ In fact, except for one case-control study,¹⁶⁸ the total data on limb reduction defects is quite limited.¹⁶⁹

164. See Jon Powell, *How to Tell the Truth With Statistics: A New Statistical Approach to Analyzing the Bendectin Epidemiological Data in the Aftermath of Daubert v. Merrell Dow Pharmaceuticals*, 31 Hous. L. REV. (forthcoming 1994) (manuscript at 58, of file with author). These studies together have an N of nearly 97,000 and yet they contain only 113 total limb reduction defects. *Id.*

165. In contingency table analyses typical of epidemiological research, the frequency of exposure in the population (the percentage of pregnant women using Bendectin) and the incidence of the effect (the frequency of limb reductions) both affect the study's power. See generally JAMES J. SCHLESSELMAN, *CASE-CONTROL STUDIES: DESIGN, CONDUCT, ANALYSIS* (1982). In cohort studies the frequency of an effect in any given study sample approximates the frequency in the population from which the sample was drawn. When the effect is very rare one needs very large samples to avoid Type II errors. In case-control studies the incidence of the effect is set at an artificially high level because researchers purposefully pick cases that exhibit the injury. Thus, when the incidence of an effect is rare in the population, case-control studies are much more powerful than cohort studies. See CARL F. CRAYNOR, *REGULATING TOXIC SUBSTANCES: A PHILOSOPHY OF SCIENCE AND THE LAW* 36 (1993); KAHN, *supra* note 116, at 54.

166. See Powell, *supra* note 164, at 58.

167. *Id.* at 61. This assumes an Alpha of .05 and a two tailed test of significance. *Id.*

168. Janet McCredie et al., *The Innocent Bystander: Doxylamine/Dicyclomine, Pyridoxine and Congenital Limb Defects*, 140 MED. J. AUSTL. 525 (1984).

169. Pooling and meta analysis combine data from several studies, producing a larger N and, *ceteris paribus*, greater statistical power. See generally Einarson et al., *supra* note 63 (presenting a step-by-step method for conducting a meta-analysis of epidemiological data); Wolf, *supra* note 63 (basic text on meta-analysis). Even these techniques cannot completely rule out the possibility that Bendectin is a weak teratogen. Powell performed meta analyses on studies with limb reduction data. See Powell, *supra* note 164. In a meta analysis of case-control studies, the Odds Ratio was 1.1, with a Chi Square of .2, $p < .65$, and a 95% confidence interval of 0.56-2.17. A meta analysis of cohort studies generated an Odds Ratio of .89, with a Chi Square of .03, $p < .86$ and a 95%

Plaintiff experts Shanna Swan and Alan Done have criticized the published studies for all of these reasons and challenged the statistical conclusion validity of their results.¹⁷⁰ They have also recoded and reanalyzed the data in some studies. Are such reanalyses so invalid that they should be excluded under Rule 702? The proffered testimony of Dr. Swan and Dr. Done supply two examples of plaintiff efforts to reanalyze the epidemiological evidence and provide a framework for addressing this question.

1. The Testimony of Dr. Swan

The first example involves Dr. Shanna Swan's reanalysis of the Center for Disease Control epidemiological data published by Cordero.¹⁷¹ A fundamental threat to the internal validity of epidemiological research is recall bias.¹⁷² Mothers bearing children with a birth defect may sift through their pre-natal experience in search of an explanation for the injury. As a consequence, in case-control studies employing a control group of healthy babies, the case mothers will remember more drug exposures which produces a biased result. Researchers can try to alleviate this threat in several ways. One alternative is to examine the prescription records of the mother's physician. Epidemiologists at the Center for Disease Control employed another

confidence interval of 0.20-4.02. A meta analysis combining all studies generated an Odds Ratio of 1.05, with a Chi Square of .06, $p < .31$, and a 95% confidence interval of 0.42-2.62. *Id.* at 78. Plaintiff experts have criticized these techniques, in part because they may treat all included studies as if they were of equal quality. See Testimony of Dr. Shanna Helen Swan, morning session, Sept. 19, 1991, at 39ff, *Havner v. Merrell-Dow Pharmaceuticals, Inc.*, No. 88-3915-F (Tex. Dist. Ct., 214th Jud. Dist., March 17, 1994) (on file with the author).

The limited number of limb defects in the epidemiological literature raises the issue of whether the relationship between Bendectin use and other types of defects is relevant to the question of whether Bendectin causes limb reduction defects. This, of course, is a question of external validity. Bendectin plaintiffs and defendants have, from time to time, been on both sides of this issue. See Sanders, *From Science to Evidence*, *supra* note 25, at 26.

170. See, e.g., Testimony of Dr. Shanna Helen Swan, morning session, Sept. 19, 1991 at 1-81, *Havner v. Merrell-Dow Pharmaceuticals, Inc.*, No. 88-3915-F (Tex. Dist. Ct., 214th Jud. Dist., March 17, 1994) (on file with the author).

171. Cordero et al., *supra* note 114, at 2307.

172. Recall bias is just one of many potential sources of bias that threaten the internal validity of epidemiological studies. Other important sources of bias are: publication bias—only studies that uncover significant results are published; and the existence of confounders that interact with the drug in question to produce injury. See Green, *supra* note 57, at 649-51; David L. Sackett, *Bias in Analytic Research*, 32 J. CHRONIC DISEASES 51, 51 (1979).

particularly innovative method in analyzing data from the Metropolitan Atlanta Congenital Defects Program.¹⁷³ The investigators divided the data into categories of birth defects and then examined the rate of first-trimester Bendectin exposure for each defect.¹⁷⁴ They compared these "cases" to a control group composed of infants with birth defects other than the one being evaluated.¹⁷⁵ Because all children in the study, both cases and controls, suffered from some birth defect, the study minimized recall bias.¹⁷⁶ As the authors noted, this technique would not allow them to detect an effect if a substance under investigation uniformly increased the risk of all types of defects investigated.¹⁷⁷ The authors discounted this possibility, however, noting that known human and animal teratogens cause specific birth defects or patterns of defects.¹⁷⁸

Dr. Swan argued that, if Bendectin causes more than one kind of birth defect, this technique would underestimate its teratogenic effects because Bendectin exposure would cause some of the control group defects and the resulting analysis would underestimate the drug's effect.¹⁷⁹ Dr. Swan re-read interview forms from the study and corrected what she perceived to be coding errors in drug use or date of exposure.¹⁸⁰ Then, in order to avoid a control group of children with defects potentially caused by Bendectin exposure, she chose as a control only those children afflicted with Down's Syndrome and other known genetic disorders.¹⁸¹ Dr. Swan reasoned that, because researchers know that drug exposure does not cause these injuries, diagnostic bias would not attenuate the results.¹⁸² Using this new control group, her reanalysis produced a significant correlation between Bendectin and limb reduction defects.¹⁸³

Measured by a validity standard, Dr. Swan's use of a different control group seems reasonable. It attempted to achieve some of what the CDC investigators hoped to achieve by using a control group comprised of children with defects while also controlling for a separate threat to internal validity arising from the

173. Cordero et al., *supra* note 114, at 2307-09.

174. *Id.* at 2307.

175. *Id.*

176. *Id.* at 2310.

177. *Id.*

178. *Id.*

179. *Lynch v. Merrell-Nat'l Lab., Inc.*, 830 F.2d 1190, 1194 (1st Cir. 1987).

180. *Id.* at 1195.

181. *Id.*

182. *Id.*

183. *Id.*

“misdiagnosis” of the control group. There is, however, an additional factor to consider. As the First Circuit observed in *Lynch v. Merrell Dow Pharmaceuticals, Inc.*, the Odds Ratio between Bendectin exposure and Down’s Syndrome in the Atlanta sample was 0.57—children suffering from Down’s Syndrome were less likely to have been exposed to Bendectin than children suffering from other defects.¹⁸⁴ Cordero’s study prominently reported this result and Dr. Swan must have known this when she chose to use children with genetic defects as the control group. A comparison between the Down’s Syndrome children and children with most other defects will produce an Odds Ratio substantially in excess of 1.0.¹⁸⁵ The First Circuit referred to this when it dismissed Swan’s analysis:

As far as appears from what is in the record, Swan made no allowance for the possibility that the very fact of having such a severe genetic deficiency as Down’s Syndrome might operate to make other rare deficiencies such as limb reduction less likely to occur in the control group — that is, that the combination of Down’s syndrome and another major misfortune might be extremely unusual. Without accounting for this possible skewing of the control group, Swan’s basis for her comparative conclusion is not apparent.¹⁸⁶

The court provided no authority to support its assertion and the Cordero Study’s data supplies no evidence for this proposition. The study does not imply that Down’s Syndrome is a prophylactic against other types of defects; that conclusion would require a comparison of Down’s Syndrome children and all other children for the existence of an additional defect. Absent some authority that Down’s Syndrome has this effect, the court’s assertion is little more than unsubstantiated hypothesis, hardly the type of validity threat that merited rejecting Dr. Swan’s testimony.

This does not mean that Dr. Swan’s analysis is preferable to that of Cordero and his colleagues. In fact, there are at least three reasons why their analysis is superior. First, Dr. Swan’s analysis creates a greater likelihood of recall bias. Because the parents of a Down’s Syndrome child know that the defect has

184. *Id.* at 1195. The study coded approximately 6% (10 of 166) of the Down’s Syndrome children as exposed to Bendectin. In the entire sample, approximately 9.5% (117 of 1,231) of the children were exposed. Cordero et al., *supra* note 114, at 2308, tbl. 2.

185. The Cordero study reported a 1.18 Odds Ratio for limb reductions. *Id.* Using only the data reported in the study and, using the Down’s Syndrome children as the controls, the Odds Ratio for limb reductions is approximately 2.3 $(14 \times 156) / (10 \times 115)$. See *id.*

186. *Lynch*, 830 F.2d at 1195.

genetic origins, they are less likely to search through their prenatal experience for possible chemical causes of the injury. In addition, the incidence of morning sickness among mothers carrying children with Down's Syndrome is unknown. If the incidence is lower than other mothers experienced, Down's Syndrome mothers would presumably be less likely to take any morning sickness medication which would make Down's Syndrome children an inappropriate control. Finally, the control group was inappropriate because Dr. Swan knew the result she would obtain before she conducted her reanalysis. Because Dr. Swan knew before she began that her comparison would produce a positive correlation between Bendectin use and limb reduction defects, the analysis could not test this hypothesis; it could not produce a negative answer. The possibility of a Type II error was zero and the possibility of a Type I error was essentially unknowable. Perhaps this lies at the heart of the First Circuit's rejection of her analysis. Although it is particularly troublesome when an investigator preparing an analysis for litigation knows *a priori* that the results will support the client's position, it does not necessarily follow that a court should exclude such testimony. The problem with Dr. Swan's testimony arose because she completed the reanalysis for the purposes of litigation; she was not testing a research hypothesis. To exclude her testimony on this ground, however, would condemn many, if not most re-analyses of existing data by experts hired for litigation.

2. The Testimony of Dr. Done

The second example is Dr. Done's proffered testimony in *DeLuca v. Merrell Dow Pharmaceuticals, Inc.*¹⁸⁷ When the case first appeared before him, Judge Brown entered summary judgment for the defendant.¹⁸⁸ He first affirmed a Magistrate's order excluding all *in vitro* and *in vivo* studies.¹⁸⁹ He then held that the plaintiff's expert testimony on epidemiology was inadmissible because it lacked the requisite Rule 703 foundation.¹⁹⁰

187. 131 F.R.D. 71 (D. N.J. 1989), *rev'd*, 911 F.2d 941 (3d Cir. 1990), *on remand*, 791 F. Supp. 1042 (D. N.J. 1992), *aff'd*, 6 F.3d 778 (3d Cir. 1993), *cert. denied*, 114 S. Ct. 691 (1994).

188. *Id.* at 74.

189. See *DeLuca v. Merrell Dow Pharmaceuticals, Inc.*, 791 F. Supp. 1042, 1045 (D. N.J. 1992), *aff'd*, 6 F.3d 778 (3d Cir. 1993), *cert. denied*, 114 S. Ct. 691 (1994). The plaintiff did not challenge this portion of the ruling. *Id.* at 1045, 1047 n.8.

190. *DeLuca*, 131 F.R.D. at 74.

The Third Circuit reversed, first noting that the primary difference between Dr. Done and the opposing epidemiologists was that Dr. Done subscribed to the approach advocated by Professor Kenneth Rothman,¹⁹¹ which deemphasizes traditional significance testing in favor of reporting relative risks and confidence intervals surrounding estimates of relative risk.¹⁹² The court held that without a record-supported, factual finding that the data Dr. Done used was not of the type reasonably relied upon by experts in epidemiology, Rule 703 did not bar his testimony.¹⁹³ Because Dr. Done used data from the same published epidemiological studies the defense relied on, the court expressed serious doubts that such a finding would be possible.¹⁹⁴ Turning to Rule 702, the court noted that the admissibility of Dr. Done's analysis was susceptible of judicial notice to the extent he based it on traditional epidemiological methodology.¹⁹⁵ Because the existing record was insufficient to make this decision, the court remanded and invited the trial judge to conduct hearings and obtain expert assistance in determining whether the evidence was sufficiently reliable to be admissible.¹⁹⁶ Finally, the Third Circuit specifically refused to decide whether epidemiological proof is inadmissible unless the data allow one to reject the null hypothesis at a .05 level of statistical significance, leaving the question for the trial court on remand.¹⁹⁷ It did note, however, that the trial court should not focus solely upon tests of significance but rather should assess "all the risks of error posed by the proffered evidence."¹⁹⁸

In many respects the Third Circuit's analysis in *DeLuca* exhibits the best understanding of validity issues of all the Bendectin opinions. The court's appreciation for the importance

191. See Kenneth Rothman, *Modern Epidemiology* (1986); Amicus Curiae Brief for Kenneth Rothman et al. in Support of Petitioners, *Daubert v. Merrell Dow*, 113 S. Ct. 2786 (1993).

192. *DeLuca v. Merrell Dow Pharmaceuticals, Inc.*, 911 F.2d 941, 946-49 (3d Cir. 1990), *on remand*, 791 F. Supp. 1042 (D. N.J. 1992), *aff'd*, 6 F.3d 778 (3d Cir. 1993), *cert. denied*, 114 S. Ct. 691 (1994).

193. *Id.* at 953.191 *Id.*

194. *Id.* at 954.

195. *Id.* at 955-56.

196. *Id.* The defendant had urged this position. *Id.* at 954. The statistical significance requirement was, of course, at the heart of the Fifth Circuit's opinion in *Brock v. Merrell Dow Pharmaceuticals, Inc.*, 874 F.2d 307, *modified*, 884 F.2d 166 (5th Cir. 1989), *cert. denied*, 494 U.S. 1046 (1990). Recall, however, that *Brock* employed a sufficiency and not an admissibility analysis. See *Brock*, 874 F.2d at 311-15.

197. *DeLuca*, 911 F.2d at 955.

198. *Id.* at 959.

of both Type I and Type II errors, and its recognition of the existence of multiple threats to validity stand in sharp contrast to the analyses in *Lynch* and *Brock*. This very understanding, however, made it difficult for the court to announce any specific admissibility guidelines and it left those issues to the trial court.¹⁹⁹

Judge Brown proceeded to hold a five day hearing followed by extensive post-hearing submissions.²⁰⁰ The parties offered written direct testimony and oral cross examination of eight expert witnesses.²⁰¹ Based on this record the judge made 120 separate findings of fact and 41 conclusions of law.²⁰²

Judge Brown first contrasted the lack of a statistically significant relationship between Bendectin ingestion and limb reduction defects in the published literature with Dr. Done's conclusion that reanalysis demonstrated a relationship.²⁰³ The parties did not dispute that Dr. Done's underlying data were of a type reasonably relied upon by experts in epidemiology.²⁰⁴ They did dispute, however, the validity of Dr. Done's calculations and the manner in which he presented his ultimate results, what Judge Brown characterized as "the methodology employed by Dr. Done."²⁰⁵ Addressing this methodology, Judge Brown noted occasions in which Dr. Done included data that he arguably should have omitted,²⁰⁶ other occasions where he excluded data that he should have included,²⁰⁷ and still other occasions where he selectively reported data.²⁰⁸ The judge noted that Dr. Done's reanalysis did not give greater weight to studies with larger number of exposed defects, the studies with the greatest power.²⁰⁹ Dr. Done also failed to weight studies based on their design or control for other sources of bias.²¹⁰ Dr. Done did not attempt to reach a quantitative conclusion based on his reevalu-

199. *DeLuca*, 791 F. Supp. at 1044.

200. *Id.*

201. *Id.* at 1044-59.

202. *Id.* at 1059.

203. *Id.* at 1045-46.

204. *Id.* at 1047 n.10.

205. *Id.*

206. *Id.* Judge Brown then proceeded to examine Dr. Done's calculations and presentation in considerable detail. He compared Dr. Done's calculations of relative risks with those of the defense experts and Dr. Shanna Swan, the plaintiff's other expert witness. On several occasions, he noted that Dr. Swan's analysis contravened Dr. Done's. *Id.* at 1047-49.

207. *Id.* at 1050.

208. *Id.* at 1051.

209. *Id.* at 1051.

210. *Id.* at 1051-52.

ation of the data, either by way of pooled data or a meta analysis.²¹¹ Finally, the Judge noted a number of ways in which Dr. Done's presentation was misleading.²¹²

The judge then proceeded to hold Dr. Done's testimony inadmissible under both Rules 702 and 703.²¹³ He followed a five part Rule 702 analysis which considered: the novelty of the technique; the existence of a specialized literature; the expert's qualifications; the non-judicial uses to which the scientific techniques are put; and the frequency with which the technique leads to erroneous results.²¹⁴ With respect to each element, Judge Brown found Dr. Done's testimony to be wanting.²¹⁵ Although the judge did not discuss scientific validity in making his 702 ruling, the opinion can be translated into this language. Statistical conclusion validity posed the biggest threat to Dr. Done's findings. The thrust of his analysis was that, although individual studies fail to produce a statistically significant relationship between Bendectin and limb reduction defects, an analysis of all the data together do reveal a relationship. Unfortunately, Dr. Done selected his 106 "data sets" in a manner that made it difficult to assess the relative likelihood of Type I and Type II errors. One gets the sense that Dr. Done engaged in a fishing expedition, sifting through large number of correlations in search of significant relationships. Moreover, the re-analyses contained measurement errors due to unreliable coding, incorrect calculations or both.

The judge also held the testimony inadmissible under Rule 703, concluding that experts in the field would not use the data Dr. Done relied on in rendering an opinion.²¹⁶ Again, the judge

211. *DeLuca*, 791 F. Supp. at 1052.

212. *Id.* at 1053. "Dr. Done's statement that 70% of his data sets have an upper confidence level above 2.0 is misleading without the corresponding information that 94% have lower confidence limits below 2.0 and only 30% of his data sets have a relative risk greater than 2.0." *Id.* In Dr. Done's analysis, a "data set" is a reported risk ratio between Bendectin use and a defect. Some studies appear several times in Dr. Done's analysis because he reports relationships for more than one type of defect. Other studies appear only once. There were a total of 106 data sets, all apparently given equal weight in constructing statements such as the one quoted above. *Id.* at 1052.

Judge Brown also noted that: "Although Dr. Done stated in his report that '92%' of the studies are compatible with an increase, he did not mention in his report that the studies were also compatible with a decrease (a proposition which he readily admits)." *Id.* at 1053 (citations omitted).

213. *Id.* at 1059.

214. *Id.* at 1056.

215. *Id.*

216. *Id.* at 1059.

presented a detailed, particularistic analysis. Epidemiological studies are, of course, of a type reasonably relied upon by experts in the field. Dr. Done's recalculations, however, produced "new data" that "has not and cannot in many instances be replicated by other experts in the field or even be explained."²¹⁷ Because Dr. Done's testimony was inadmissible under Rules 702 and 703, the plaintiffs could not meet their burden of proof on the issue of causality and the court granted the defendant's summary judgment motion.²¹⁸

The testimony Dr. Done proffered in *DeLuca* presents a particularly difficult case. From a validity point of view it is tempting to agree with the judge's conclusion. At several points the judge implied that the basic flaw in Done's analyses was that neither defense nor plaintiff experts were able to replicate a number of Dr. Done's conclusions.²¹⁹ This inability to replicate the data, however, does not present a problem if one can trace Dr. Done's methods with sufficient specificity to attempt a replication. Under these circumstances Dr. Done's conclusions would be falsifiable and, in this core sense, scientific.²²⁰ Because a substantial number of Dr. Done's conclusions were falsifiable, however, Judge Brown was able to demonstrate how they were in error. Indeed, it is this detailed demonstration of error that gives power to the opinion. The court did not question epidemiology as a valid methodology. Rather, the judge challenged Dr. Done's particular recalculations and the "new data" these suspect recalculations generated. At this level of analysis the court found Dr. Done's testimony to be flawed largely because it was not a neutral rendition of the epidemiological evidence concerning Bendectin. The judge depicted Dr. Done as a "party witness" who designed his analysis to advance his employer's case. Less apparent, however, is whether Dr. Done's testimony was significantly different from that of experts in other cases and whether this level of analysis casts doubt on the admissibility of a great deal of expert testimony.

This question is unlikely to be answered any time soon. Although the Third Circuit demonstrated particular concern for the dangers of devising a special rule for Bendectin cases, Judge Brown did just this on remand. The judge noted that ordinarily the inclusion and exclusion of certain data is a matter for the

217. *DeLuca*, 791 F. Supp. at 1059.

218. *Id.*

219. *Id.* at 1048, 1059.

220. See Black et al., *supra* note 17, at 68-70.

battle of the experts but in this case, Dr. Done's plaintiff-leaning, combined with the errors and uncertainties in his calculations, threatened to confuse the jury.²²¹ It is hard to imagine that this consideration would carry as much weight if the scientific evidence is not fully developed and, therefore, the errors in the expert's analysis are not so obvious. Likewise, the importance the trial court placed on the fact that Dr. Done had not published his work in a peer review journal was inevitably influenced by the existence of a large body of published research on Bendectin. The lack of publication would not weigh so heavily if there were very little published research.²²² Most important, such a detailed analysis of an expert's proffered testimony is unlikely to occur in more than that handful of cases in which the courts have become particularly concerned with issues of judicial efficiency and jury confusion.

3. Summary

The rejection of epidemiological evidence in *Lynch* and *DeLuca* again indicates the uncertain nature of admissibility determinations under a *Daubert*-like analysis. The proffered testimony of both Drs. Done and Swan presented validity problems. Scientific validity, however, encompasses a complex set of concepts and is always a matter of degree. It is difficult to pinpoint exactly why the courts found that this testimony was so invalid as to be inadmissible. Dr. Done's situation is particularly instructive. Even were one to conclude that the trial judge's opinion fairly reflects Dr. Done's proffered testimony, this at best supports excluding 50%, 60% or perhaps even 80% of his testimony on scientific validity grounds. It does not, however, fairly support the exclusion of all Dr. Done's testimony. Importantly,

221. *DeLuca*, 791 F. Supp. at 1058.

222. Most debate over the peer review process concerns its ability to monitor the scientific validity of reported findings. See, e.g., Thomas S. Burack, *Of Reliable Science: Scientific Peer Review, Federal Regulatory Agencies, and the Courts*, 7 VA. J. NAT. RESOURCES L. 27 (1987). In this regard, the value of peer review is frequently overrated, as any academic who has been a reviewer can attest. Less frequently noted is that publication in a reputable peer review journal imposes a style of discourse that encourages a relatively conservative, dispassionate, and neutral presentation rarely found in trial testimony. Peer reviewed, published articles are less likely to overstate or understate the value of a particular finding or use causal language to describe their results. See Dan L. Burk, *When Scientists Act Like Lawyers: The Problem of Adversary Science*, 33 JURIMETRICS J. 363, 368 (1993); Robert Rosenthal & Peter David Blanck, *Science and Ethics in Conducting, Analyzing, and Reporting Social Science Research: Implications for Social Scientists, Judges, and Lawyers*, 68 IND. L.J. 1209, 1212 (1993).

neither *Lynch* nor *DeLuca* made clear why the problems with the proffered testimony did not go to weight rather than admissibility. A full understanding of these opinions lies beyond questions of scientific validity and even beyond questions of admissibility.

V. REASONS FOR RESTRICTIVE ADMISSIBILITY RULINGS

Why have courts been so willing to make restrictive admissibility rulings in *Bendectin* cases? In part, the answer can be found in the objectives courts attempt to achieve by restricting the scope of admissible scientific testimony and in the special problems mass torts pose. There are at least two reasons to restrict the admissibility of scientific evidence. First, restricting this evidence fosters judicial efficiency. If a party's scientific arguments are without merit, excluding them minimizes the expenditure of resources required to resolve the issue and husbands scarce judicial resources for the resolution of closer questions.²²³ A second reason to restrict this testimony is that juries²²⁴ will likely be unable to distinguish between reliable and unreliable evidence.²²⁵ Although juries may be good factfinders with respect to lay testimony, some argue that their lack of specialized knowledge renders them incapable of assessing the merits of expert testimony.²²⁶ Restrictions on admissibility reduce the probability that "a credulous jury will now and again transform scientific dust into gold."²²⁷

223. See Sanders, *The Bendectin Litigation*, *supra* note 24, at 301.

224. On the question of whether the judge or jury is a better factfinder, see Phoebe C. Ellsworth, *Are Twelve Heads Better Than One?*, 52 *LAW & CONTEMP. PROBS.* 205, 217-18 (1989); Richard Lempert, *Civil Juries and Complex Cases: Taking Stock after Twelve Years*, in *VERDICT: ASSESSING THE CIVIL JURY SYSTEM* 181 (Robert Litan ed., 1993); Robert MacCoun, *Inside the Black Box What Empirical Research Tells Us About Decisionmaking by Civil Juries*, in *VERDICT: ASSESSING THE CIVIL JURY SYSTEM* 137 (Robert Litan ed., 1993); Sanders, *From Science to Evidence*, *supra* note 25, at 82.

225. "The principal argument for reviewing expert testimony is the concern over jurors' ability to discount unreliable expert testimony appropriately." Faigman, *supra* note 52, at 881.

226. See generally HUBER, *GALILEO'S REVENGE*, *supra* note 13.

227. *Id.*; see *THE EVOLVING ROLE OF STATISTICAL ASSESSMENTS AS EVIDENCE IN THE COURTS* 150 (Stephen E. Fienberg ed., 1989) [hereinafter *THE EVOLVING ROLE*]. In *Turpin v. Merrell Dow Pharmaceuticals, Inc.*, the court cited jury confusion as one of the reasons for closely reviewing scientific evidence. 959 F.2d 1349, 1352 (6th Cir. 1992), *cert. denied*, 113 S. Ct. 84 (1992).

These considerations apply with special force to mass torts, such as the Bendectin litigation. The many congregations²²⁸ of substance-related mass torts²²⁹ that have emerged over the last decade and a half are a new phenomenon of the tort system. The size and scope of these cases have placed enormous pressures on the judicial process and judges have reacted by seeking out new, efficient ways to dispose of them. A first step toward this goal is procedural rationing. Courts have used class actions under Rule 23 of the Federal Rules of Civil Procedure,²³⁰ consolidation under the Multi-District Litigation Act,²³¹ and consolidation of cases for trial under Rule 42 of the Federal Rules of Civil Procedure²³² in the Bendectin cases and other mass torts. Courts have likewise engaged in substantive rationing.²³³ In situations such as asbestos exposure, the courts, reasonably confident that

228. On the concept of case congregations, see Marc Galanter, *Case Congregations and Their Careers*, 24 LAW & SOC'Y REV. 371 (1990); Sanders, *The Bendectin Litigation*, *supra* note 24, at 307.

229. Substance-related mass torts should be distinguished from mass torts generated by a single event such as an airplane crash.

230. FED. R. CIV. P. 23; see Richard A. Chesley & Kathleen Woods Kolodgy, *Mass Exposure Torts: An Efficient Solution to a Complex Problem*, 54 U. CIN. L. REV. 467, 490-96 (1985); Linda S. Mullenix, *Class Resolution of the Mass-Tort Case: A Proposed Federal Procedure Act*, 64 TEX. L. REV. 1039, 1043 (1986).

231. Multi-District Litigation Act, 28 U.S.C. § 1407 (1988). For discussions of consolidation of multi-district litigation, see DAVID F. HERR, MULTIDISTRICT LITIGATION: HANDLING CASES BEFORE THE JUDICIAL PANEL ON MULTIDISTRICT LITIGATION (1986); Cornelius J. Moynihan, Jr., *Multiple Products Liability Suits and Their Collateral Estoppel Aspects*, 73 MASS. L. REV. 83 (1988); Blake M. Rhodes, *The Judicial Panel on Multidistrict Litigation: Time For Rethinking*, 140 U. PA. L. REV. 711 (1991); Joan Steinman, *Law of the Case: A Judicial Puzzle in Consolidated and Transferred Cases and in Multidistrict Litigation*, 135 U. PA. L. REV. 595, 622-26 (1987).

232. FED. R. CIV. P. 42. See generally Linda S. Mullenix, *Beyond Consolidation: Postaggregative Procedure in Asbestos Mass Tort Litigation*, 32 WM. & MARY L. REV. 475, 500 (1991).

233. For a thoughtful early discussion of these and other devices designed to deal with mass torts, see Jack B. Weinstein, *Preliminary Reflections on the Law's Reaction to Disasters*, 11 COLUM. J. ENVTL. L. 1 (1986). Courts continue to explore different means of rationing law in mass tort cases. Recent initiatives include attempts to create mandatory limited fund class actions, (see *In re Joint Eastern and Southern Dist. Asbestos Litig.*, 14 F.3d 726, 728 (2d Cir. 1993)), the settlement of claims of future plaintiffs, (see *Carlough v. Amchem Products, Inc.*, 5 F.3d 707, 710 (3d Cir. 1993); *\$4.75 Billion Settlement Proposed for Silicone Breast Implant Cases*, Prod. Safety & Liab. Rep. (BNA) 1 (September 14, 1993) (settlement of all existing and future breast implant cases)) and the use of the All Writs Act to prohibit some claimants from bringing claims in state courts (see *In Re "Agent Orange" Product Liability Litig.*, 996 F.2d 1425, 1431 (2d Cir. 1993), *cert. denied*, 114 S. Ct. 1125 (1994)). All these means share a common feature: they deny individual claimants the right to a separate, individualized trial of their cause of action.

many plaintiffs have valid claims, have frustrated defendants' efforts to tie up the legal system by relitigating previously tried issues²³⁴ or demanding a separate trial for each claimant.²³⁵ Conversely, in the Bendectin cases, the courts became increasingly certain that plaintiffs did not have valid claims and sought to prevent separate trials for those individual plaintiffs who had not litigated their claims in a 1985 consolidated trial in the Southern District of Ohio.²³⁶ Admissibility rulings provided one of the few devices available to achieve the goal of non-suiting these plaintiffs.

Likewise, concern about the jury's ability to understand scientific evidence is particularly salient in mass tort cases. Here, as in any lawsuit where multiple juries try similar facts, an inability to understand the evidence may produce inconsistent results. In most situations inconsistent verdicts can be explained in terms of unique facts presented in one case and not another. In mass tort cases, however, the outcome frequently turns on questions of general causation and inconsistencies are not easily hidden.²³⁷ Concerns about jury inconsistency echoes through

234. Offensive collateral estoppel has proven to be an unsuccessful device to achieve the goal of improving efficiency in mass tort cases. Michael D. Green, *The Inability of Offensive Collateral Estoppel to Fulfill It's Promise: An Examination of Estoppel in Asbestos Litigation*, 70 IOWA L. REV. 141, 186-87, 207-12 (1984) [hereinafter Green, *Offensive Collateral*]; see *Setter v. A.H. Robins Co.*, 748 F.2d 1328, 1330 (8th Cir. 1984); *Hardy v. Johns-Manville Sales Corp.*, 681 F.2d 334, 340-41 (5th Cir. 1982). Other tactics have been more successful. In *Beshada v. Johns-Manville Products Corp.*, the New Jersey Supreme Court ruled that the defendant could not raise a state-of-the-art defense in an asbestos products liability action. 447 A.2d 539, 542 (1982). Although the court quickly retreated from this position in *Feldman v. Lederle Laboratories*, it did not overturn *Beshada*, but restricted it to "the circumstances giving rise to its holding." 479 A.2d 374, 387-88 (1984). One interpretation of this statement is that, after a substantial amount of asbestos litigation, the *Beshada* court engaged in substantive rationing by refusing to allow the defense to use an argument that had failed repeatedly in the past. See *In Re Asbestos Litigation*, 829 F.2d 1233, 1243-44 (3d Cir. 1987), *cert. denied*, 485 U.S. 1029 (1988).

235. See *Cimino v. Raymark Indus., Inc.*, 751 F. Supp. 649, 651-52, 662 (E.D. Tex. 1990) (rejecting an individual-by-individual approach to damage awards in a class action suit); Deborah Hensler, *Resolving Mass Toxic Torts: Myths and Realities*, 1989 U. ILL. L. REV. 89, 90, 103-04 (1989) (discussing the development of aggregation in mass claims).

236. *In Re Richardson-Merrell, Inc. "Bendectin" Products Liability Litig.*, 624 F. Supp. 1212, 1250 (S.D. Ohio, 1985), *aff'd sub nom.*, *In Re Bendectin Litig.*, 857 F.2d 290 (6th Cir. 1988), *cert. denied*, 488 U.S. 1006 (1989).

237. See, e.g., Green, *Offensive Collateral*, *supra* note 234, at 215-20 (discussing concerns of inconsistent jury verdicts). Professor Green gives an example of jury inconsistency in a mass tort case where five separate juries heard the same evidence on questions common to five asbestos cases tried simultaneously in the same courtroom against twelve defendants in 1982. *Id.* at 221-22. In

the Bendectin cases, including *Lynch*²³⁸ and *Brock*.²³⁹ Inconsistency may, of course, reflect the fact that a trial poses a particularly close fact pattern and that reasonable juries may differ about the correct outcome. If so, over time jury verdicts should produce outcomes that reflect the underlying strength of the parties' case.²⁴⁰ The verdicts will define the expected value of a case and, if that value is low enough, past verdicts will deter future claimants.²⁴¹ Faced with this type of inconsistency, courts might wish to facilitate settlements but would not necessarily want to interfere with the trial process.

Inconsistency may reflect a more fundamental problem, however: that an unacceptable percentage of juries are reaching "incorrect" verdicts and, therefore, the verdicts as a group do not reflect the merits of the issue. Some Bendectin opinions evidence a belief that jury verdicts for the plaintiff were erroneous.²⁴² The restrictive admissibility opinions reflect a judicial

response to special interrogatories the juries disagreed about whether some or all of the products were defectively designed and marketed; whether asbestos exposure was the sole cause of mesothelioma; whether the defendant, Johns-Manville, was grossly negligent; and the date on which the defendant should have foreseen the dangers associated with work-place asbestos exposure. *Id.* at 222. As to this latter determination, the jury answers ranged from 1935 to 1965. *Id.* at 222-23, 228-35.

238. *Lynch v. Merrell-Natl' Lab.*, 646 F. Supp. 856, 861-62 (D. Mass. 1986), *aff'd*, 830 F.2d 1190 (1st Cir. 1987).

239. *Brock v. Merrell Dow Pharmaceuticals, Inc.*, 874 F.2d 307, 310 (5th Cir.) (noting that inconsistent jury verdicts suggest that appellate courts should resolve such questions), *modified*, 884 F.2d 166 (5th Cir. 1989), *cert. denied*, 494 U.S. 1064 (1990); *see Turpin v. Merrell Dow Pharmaceuticals, Inc.*, 959 F.2d 1349, 1349 (6th Cir. 1992), *cert. denied*, 113 S. Ct. 84 (1992); *DeLuca v. Merrell Dow Pharmaceuticals, Inc.*, 911 F.2d 941, 951-52 (3d Cir. 1990), *aff'd*, 6 F.3d 778 (3d Cir. 1993), *cert. denied*, 114 S. Ct. 691 (1994).

240. A question related to verdict consistency is damages consistency. Once the defendant's liability has been adjudicated or conceded, one intriguing solution to damage inconsistency is to average damage awards from a representative sample of cases chosen for trial and apply this result to untried cases. The court in *Cimino v. Raymark Industries* adopted this approach. 751 F. Supp. 649, 664-65 (E.D. Tex. 1990). This solution greatly reduces transaction costs, produces a better estimate of the plaintiffs' "true" damages than any individual verdict, and promotes fairness between different plaintiffs. *See* Glen O. Robinson & Kenneth S. Abraham, *Collective Justice in Tort Law*, 78 VA. L. REV. 1481, 1490-96 (1992); Michael J. Saks & Peter David Blanck, *Justice Improved: The Unrecognized Benefits of Aggregation and Sampling in the Trial of Mass Torts*, 44 STAN. L. REV. 815, 815 (1992).

241. *See Galanter*, *supra* note 228, at 388-93; Francis McGovern, *Toward a Functional Approach for Managing Complex Litigation*, 53 U. CHI. L. REV. 440, 478-83 (1986).

242. In *Brock v. Merrell Dow Pharmaceuticals, Inc.*, for example, the court was concerned that inconsistent verdicts would over-deter defendants and thus hinder the development of new drugs. 874 F.2d 307, 310, *modified*, 884 F.2d

belief that, left alone, too many juries will reach an incorrect outcome. Judges²⁴³ and others²⁴⁴ are particularly suspicious of the jury's ability to arrive at a correct decision in trials involving the expert presentation of complex technological and scientific questions, trials that typify much of mass tort litigation. The outcome of the Bendectin trials do little to alleviate this concern. Of the twenty jury trials that have reached a verdict on the merits, eight resulted in a plaintiff victory.²⁴⁵ This 40% success rate mirrors the overall success rate in product liability cases.²⁴⁶ Thus, the one-sided nature of the scientific evidence has not resulted in a perceptible tilt in favor of Bendectin defendants. Restrictive admissibility determinations may be interpreted as a response to these "incorrect" verdicts. They are, from this point of view, an ad hoc method of jury control.

The twin objectives of achieving an efficient use of judicial resources in mass tort cases and assisting the jury in understanding scientific evidence are meritorious goals. Restrictive admissibility rulings, however, are a flawed means to these objectives. With respect to the efficiency goal, the courts have attempted to achieve with admissibility rulings what they should be achieving with sufficiency rulings. Indeed, several of the Bendectin opinions employed a sufficiency analysis. The *Brock* court had a relatively easy task because it had a full transcript on which to rule. The courts in *Lynch* and *DeLuca* did not enjoy this luxury. Instead, they encountered expert witnesses prepared to testify that Bendectin more likely than not caused the plaintiff's injury. If these courts admitted even part of the proffered testimony, the plaintiff would be able to present a prima facie case on causation, making summary judgment for the defendant inappropriate. Therefore, excluding all of the plaintiff's causation evidence was a necessary prerequisite to ruling for the defendant as a matter of law. In every Bendectin opinion that excluded the plaintiff's expert testimony on causation, the court ultimately entered a judgment for the defendant.

166 (5th Cir. 1989), cert. denied, 494 U.S. 1046 (1990); see Alan Golanski, *Judicial Scrutiny of Expert Testimony in Environmental Tort Litigation*, 9 PACE ENVTL. L. REV. 399, 465 (1992).

243. See *In re Japanese Electronic Products Antitrust Litig.*, 631 F.2d 1069, 1084-85 (3d Cir. 1980) (due process considerations may create a complex case exception to the right to a jury trial).

244. William V. Luneberg & Mark Nordenberg, *Specially Qualified Juries and Expert Nonjury Tribunals: Alternatives for Coping with the Complexities of Modern Civil Litigation*, 67 VA. L. REV. 887 (1981); see Lempert, *supra* note 224.

245. Sanders, *From Science to Evidence*, *supra* note 25, at 9.

246. *Id.* at 5 n.16.

This need to find all of the plaintiff's causation testimony inadmissible helps to explain the restrictive rulings in the Bendectin cases. The courts, persuaded that the plaintiffs had insufficient evidence to prevail on the merits, sought to avoid still another trial that might result in a verdict for the plaintiff and require them to enter a j.n.o.v., as in *Richardson*,²⁴⁷ *Brock*,²⁴⁸ and *Ealy*.²⁴⁹ Efficiency considerations, therefore, play an important role in these rulings.

Using admissibility rulings in this way has several drawbacks. The Bendectin admissibility decisions confuse an already blurry line between admissibility and sufficiency.²⁵⁰ Sufficiency necessarily entails a decision about the entire body of the party's case. Perhaps in a *Frye*-world admissibility might be thought to raise a similar question: whether a significant part of the scientific community believes that the case presents an arguable scientific issue. After *Daubert*, however, admissibility is best described as a decision about individual pieces of scientific evidence, a conception ill-suited to the global assessment of the science supporting a party's position.

Moreover, as others have observed,²⁵¹ the Bendectin cases are unique in a number of ways including the existence of an unusually rich body of epidemiological data, an extensive legal record produced by hundreds of cases and thirty trials and, perhaps most fundamentally, a relatively one sided body of scientific evidence. Admissibility criteria created to dispense with Bendectin cases may present problems in other areas where the science is neither as well developed nor its weight as one sided.²⁵² Perhaps this is inevitable when courts bend the rules of evidence to foster efficiency. If 80% of a party's expert testimony can be excluded, a court will be very reluctant to admit the last 20% and allow a trial on the merits. The court will find reasons to exclude the remaining 20% which may undermine a so-

247. *Richardson v. Richardson-Merrell, Inc.*, 857 F.2d 823 (D.C. Cir. 1988), cert. denied, 493 U.S. 882 (1989).

248. *Brock v. Merrell Dow Pharmaceuticals, Inc.*, 874 F.2d 307 (5th Cir.), modified, 884 F.2d 166 (5th Cir. 1989), cert. denied, 494 U.S. 1046 (1990).

249. *Ealy v. Richardson-Merrell, Inc.*, 897 F.2d 159 (D.C. Cir.), cert. denied, 498 U.S. 950 (1990).

250. See Susan R. Poulter, *Science and Toxic Torts: Is there a Rational Solution to the Problem of Causation?*, 7 HIGH TECH. L.J. 189, 193-94 (1992).

251. Green, *supra* note 57, at 677.

252. This risk exists for incautious sufficiency decisions as well. The best example of this is *Brock's* requirement that plaintiffs present statistically significant epidemiological evidence of a relationship between Bendectin use and their injury. *Brock*, 874 F.2d at 313-15.

phisticated approach to the question of scientific validity. Moreover, admissibility decisions that require a hearing and briefing as extensive as Judge Brown's in *DeLuca* erode efficiency gains.²⁵³

Although this Article opposes the use of admissibility rulings to non-suit plaintiffs in order to achieve efficiency goals, it does recognize that mature congregations, such as the *Bendectin* cases, do pose special problems for the courts and society. Special solutions should be developed but they should focus on the problem at hand: the repeated litigation of the same issue in a mature congregation of cases.²⁵⁴ Along these lines, Professor Berger has suggested several ways to deal with this issue.²⁵⁵

VI. THE JUROR'S PROBLEM AND ALTERNATIVE SOLUTIONS

Restrictive admissibility rules are also an inappropriate solution to the problems juries have with complex scientific arguments in mass tort cases. A problem does exist, however. A mounting body of evidence supports the position that jurors do have a difficult time understanding and assessing expert scientific testimony. For example, the American Bar Association Section on Litigation commissioned a Special Committee to study jury comprehension in complex cases.²⁵⁶ The Committee studied four complex cases in the areas of sexual harassment, anti-trust, arson-related insurance fraud, and misappropriation of trade secrets. The scientific evidence was particularly difficult in the trade secrets case²⁵⁷ and jurors reported that they had

253. Although there may be relatively few efficiency gains in a particular case, especially where *DeLuca*-like hearings are required, the restrictive admissibility rulings may have a chilling effect across the entire congregation of cases and cause plaintiffs to postpone or forego litigation that enjoys a slim chance of success.

254. One possibility would be to define certain bodies of knowledge as "social framework" information and allow the court to instruct the jury on this framework as the court instructs the jury on the "legal framework," it should use in deciding the case. See Laurens Walker & John Monahan, *Social Frameworks: A New Use of Social Science in Law*, 73 VA. L. REV. 559 (1987).

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

256. SPECIAL COMMITTEE ON JURY COMPREHENSION, JURY COMPREHENSION IN COMPLEX CASES i-ii (1989) [hereinafter JURY COMPREHENSION]. The Committee engaged Elizabeth Loftus, Jane Goodman and Edith Green to conduct the study. *Id.*

257. This judgment is based on Richard O. Lempert's article, *Civil Juries and Complex Cases: Taking Stock after Twelve Years*. See Lempert, *supra* note 224. Lempert examined thirteen complex cases and rated each on a three point

trouble understanding the facts. The authors of the report concluded the following:

[A]lthough one plaintiffs' attorney suggested that jurors would not need to understand the chemical processes in dispute to decide this case, it seems apparent that some ability to comprehend and evaluate the technical information was in fact, imperative.

It is not clear that jurors—even those employed as engineers who had completed college courses in chemistry—had that ability. Even these jurors felt overwhelmed by the technical nature of the evidence. Less educated jurors suggested that they were completely “out of their league.”²⁵⁸

The jury also had difficulty applying the facts to the jury instructions to determine whether the facts established the claims.²⁵⁹

The fact that 40% of the juries that reached the merits of the Bendectin cases found for the plaintiff is not inconsistent with the jury's experience in the trade secret case. Interviews with jurors in one Bendectin trial indicated that they also had difficulty understanding the scientific evidence.²⁶⁰ A detailed analysis of six Bendectin trial transcripts indicated some of the reasons jurors have difficulty with the scientific evidence. Most importantly, the trial structure itself makes it very difficult to weigh evidence and, at least within the tort context, to separate the scientific analysis of causation from other elements of the tort.²⁶¹ Many factors contribute to produce this result.²⁶² The parties typically employ experts whose objectivity is therefore suspect. Although both plaintiff and defense experts may testify concerning exactly the same scientific studies and findings, the structure of trials separates their testimony, sometimes by many days. Frequently, a roughly equal number of experts from each side testify on each scientific issue, producing a perception that real conflict exists within the scientific community on nearly all questions.²⁶³ Because the parties focus on the science that they believe best supports their position, the jury is likely to

difficulty scale: low, moderate, and high. *Id.* at 185-90. The trade secret case and two others were scored high. *Id.* at 185-90, tbl. 6.1.

258. JURY COMPREHENSION, *supra* note 256, at 103.

259. *Id.* at 54. Jurors in other complex cases have similar problems when the science is difficult to understand. See Sanders, *Jury Deliberation*, *supra* note 101, at 49-51.

260. *Id.* at 45.

261. Sanders, *From Science to Evidence*, *supra* note 25, at 61.

262. For a valuable discussion of the shortcomings of present methods of introducing expert testimony, see Samuel Gross, *Expert Evidence*, 1991 Wis. L. Rev. 1113 (1991).

263. Sanders, *From Science to Evidence*, *supra* note 25, at 40.

conclude that all types of scientific evidence are equally probative to the issue in dispute.²⁶⁴

Ironically, the problem courts have attempted to correct through restrictive admissibility rulings is, in part, a product of the Rules of Evidence themselves. For example, under Rule 703 experts may base their opinion on facts or data of a type reasonably relied upon by experts in the particular field in forming opinions on a subject. Any published articles the expert relied on would ordinarily be hearsay unless admitted under some exception such as the Learned Treatises exception. Even when admitted under this exception, however, "the statements may be read into evidence but may not be received as exhibits."²⁶⁵ As a consequence, the jury does not have the ability independently to examine and assess statements those works contain. Rather, it must view them through the filter of party advocacy, a filter that makes it very difficult to assess the weight of scientific opinion on an issue.

All of these difficulties reflect a more general problem with the presentation of scientific evidence. Much of what goes on at trial in America is a process of deconstructing science. As Peter Schuck has noted, law and science are in some ways competing cultures, each with its own set of central values, incentives, techniques, biases and orientations.²⁶⁶ Although science and law share a wide range of cultural values, differences do exist. The core values of these two cultures reflect these differences. Whereas science's central value is truth, law's central value, at least in its judicial manifestation, is justice.²⁶⁷ Law does not pursue justice as an abstract ideal, however, but in a context that acknowledges the existence of competing views of what constitutes a just outcome and allows these competing views to contend for supremacy within an adversarial trial. The differences between the two cultures create tensions and contests for domi-

264. For example, several jurors interviewed from the *Havner* Benedictin trial perceived the epidemiological evidence to be no more probative than animal studies or *in vitro* studies on the question of whether the drug is a teratogen. Sanders, *Jury Deliberation*, *supra* note 101, at 62.

265. FED. R. EVID. 803(18).

266. Peter Schuck, *Multi-Culturalism Redux: Science, Law and Politics*, 11 YALE L. & POL'Y REV. 1 (1993) (discussing the differing cultures of science, law and politics).

267. *Id.* at 21; see Sheila Jasanoff, *What Judges Should Know About the Sociology of Science*, 32 JURIMETRICS J. 345, 354 (1992) (making the same distinction).

nance.²⁶⁸ In the courtroom, the battle frequently involves attacking the scientific culture itself by focusing on its biases, its implicit assumptions, and the many ways it inevitably fails to live up to its own ideals of rigorous methodology and objectivity.²⁶⁹

As Sheila Jasanoff has noted, scientific discoveries, like other types of knowledge, are premised on underlying assumptions and conventions that remain in the background until controversy erupts.²⁷⁰ The assumptions include both experimental and interpretative conventions. In ordinary scientific conversations these assumptions lie in the background and scientists speak of things as being true or false. Because scientific knowledge, like other forms of knowledge, is constructed by a community of individuals, it can be deconstructed, pulled apart by questioning each assumption, each shared understanding, and each indeterminacy that inevitably infects even the most artfully crafted research. Indeed, many of the admissibility battles discussed in this Article reflect exactly this type of deconstruction and it is important to note that both plaintiffs and defendants, motivated by a lawsuit, actively participate in this process. The adversarial trial is particularly well suited to such an effort, which, in this context, is its greatest weakness:

Adversarial process is indeed a wonderful instrument for deconstructing "facts," for exposing contingencies and hidden assumptions that underlie scientific claims, and thereby preventing an uncritical acceptance of alleged truths. The adversary process is much less effective, however, in reconstructing the communally held beliefs that reasonably pass for truth in science. Cross-examination, in particular, unduly privileges skepticism over consensus. It skews the picture of science that is presented to the legal factfinder and created an impression of conflict even where little or no disagreement exists in practice.²⁷¹

From this perspective the problem confronting the courts involves more than simply assessing the validity of a particular fact, method, or conclusion. Rather, the court must find a balance between the pursuit of justice in an environment of adver-

268. [E]xpert testimony can be perceived to be a challenge to certain fundamental concepts implicit in the structure of trials in the common law tradition, which calls into question whether the common law mode of trial prevalent in the United States can accommodate expert testimony without substantial change.

Ronald J. Allen & Joseph S. Miller, *The Common Law Theory of Experts: Deference or Education?*, 87 Nw. U. L. REV. 1131, 1131 (1993).

269. Schuck, *supra* note 266, at 18.

270. Jasanoff, *supra* note 267, at 347-48.

271. *Id.* at 353-54.

sarial legalism and respect for science's culture, values and assumptions.

If the heart of the problem confronting juries is that the law so successfully deconstructs scientific findings that juries find it very difficult to assess the relative merits of any position and ultimately begin to discount the value of scientific "truth" for the resolution of the problem posed to them,²⁷² restricting the evidence they hear is a problematic cure. This cure runs the inevitable risk of balancing the excessively skeptical environment that ordinary methods create with an uncritical determination that there is good science and bad science, that the two can be distinguished, that the trial judge can capably make this distinction, and that the court can protect the jury by the excluding the bad.

Undoubtedly, admissibility decisions have a role to play in excluding marginal "science" from the courtroom.²⁷³ There are, however, superior alternatives available in most situations. These alternatives directly confront the problems that arise from the undervaluation of "normal" scientific understandings when litigants introduce science through traditional adversarial processes.²⁷⁴ They include the use of court-appointed experts and expert panels, the bifurcation of trials in order to try causal questions separate from breach of duty questions, and the poten-

272. The Special Committee of the ABA Section of Litigation concluded that jurors are not too impressed with experts and dismiss many of them as hired guns. JURY COMPREHENSION, *supra* note 256, at 40. Other research reflects similar attitudes. See Neil Vidmar, *Assessing the Impact of Statistical Evidence, A Social Science Perspective*, in THE EVOLVING ROLE, *supra* note 227, at 296-97. The perception that experts overwhelm jurors simply because they are experts is unfounded. As a juror in an asbestos case reported: "The expert testimony was not a real factor in our decision, except in the very backhanded sense that it lent medical credence to any result." Jane Goodman et al., *What Confuses Jurors in Complex Cases*, TRIAL, Nov. 1985, at 65, 68.

273. Jasanoff has noted that much, if not all, of what passes for clinical ecology may be excluded because it violates basic canons of science. Jasanoff, *supra* note 267, at 355.

274. Of course, the term "normal" is itself difficult to describe. At its core, however, may reside the idea of an "empiricist repertoire," the conversations scientists hold when they are not attacking each other's accounts of reality. Jasanoff, *supra* note 267, at 348. See generally G. NIGEL GILBERT & MICHAEL J. MULKAY, *OPENING PANDORA'S BOX: A SOCIOLOGICAL ANALYSIS OF SCIENTIST'S DISCOURSE* (1984). "Normal" science may occasionally cease to exist in an area where science itself becomes so politicized that the scientific community divides into camps that constantly attack the other side's account of reality. Revealingly, this sometimes occurs when the law "captures" an area of science and uses it to resolve very controversial disputes. See ELEANOR P. WOLF, TRIAL AND ERROR: THE DETROIT SCHOOL SEGREGATION CASE 335 n.34 (1981).

tial use of specialized science courts and blue ribbon juries.²⁷⁵ Each of these alternatives would either reduce the perception that science is mostly conflict with little consensus or reduce the degree to which the central scientific value of truth is pitted against the core legal value of justice.²⁷⁶

CONCLUSION

Mass torts continue their relentless assault on our common law tort system. They have forced us to rethink accepted methods of proving causation, the appropriate measure of damages and even the system's fundamental commitment to an individualized trial of each plaintiff's case.²⁷⁷ Many mass torts create a pair of problems for the court system. They consume judicial resources and pose complex scientific questions that frequently confuse juries. The increased use of expert scientific witnesses that has accompanied the rise of mass torts has prompted charges that many of these witnesses are introducing "junk science." The Supreme Court crafted its decision in *Daubert* against this background. From a narrow perspective, *Daubert* simply resolved a longstanding issue in the law of evidence by holding that the Federal Rules of Evidence superseded *Frye*. From a wider perspective, the opinion represents an attempt to define, or perhaps redefine, the relationship between science and the law.

The American legal system, both in its judicial and regulatory capacities, has constructed a set of structures that facilitate attacks on science and undermine trust in the judgment of the scientific community.²⁷⁸ In one sense *Daubert* attempted to redefine this relationship. By placing the concept of scientific

275. See Sanders, *From Science to Evidence*, *supra* note 25, at 67-82.

276. Specifically, the bifurcation of trials serves this end. It minimizes the parties' ability to construct presentations that invite the factfinder to trade a weak case of causation off against a stronger case of negligence. Sanders, *From Science to Evidence*, *supra* note 25, at 52. Some oppose bifurcation because it promotes "truth" over "justice." See *In re Beverly Hills Fire Litigation*, 695 F.2d 207, 217 (6th Cir. 1982), *cert. denied*, 461 U.S. 929 (1983); Roger Trangsrud, *Mass Trials in Mass Tort Cases: A Dissent*, 1989 U. ILL. L. REV. 69, 80-82 (1989).

277. Robert L. Rabin, *Some Thoughts on the Efficacy of a Mass Toxics Administrative Compensation Scheme*, 52 MD. L. REV. 951 (1993); Robinson & Abraham, *supra* note 240, at 1481-83.

278. See Sheila Jasanoff, *Acceptable Evidence in a Pluralistic Society*, in ACCEPTABLE EVIDENCE: SCIENCE AND VALUES IN RISK MANAGEMENT 29 (Deborah G. Mayo & Rachelle D. Hollander eds., 1991) (comparing British and American regulatory approaches to risk).

validity at the center of admissibility decisions, *Daubert* invoked scientific understandings of what constitutes good and bad science. The Court recognized that for science to be useful the law must attend to more than the scientist's conclusion. The legal system must, to some degree, be attentive to the scientific method itself. It must interpret scientific conclusions in the context of the methods and culture which precipitated them.

This very change, however, reveals the degree to which many restrictive rulings in the Bendectin cases cannot be justified from the perspective of scientific validity. These opinions are better explained in terms of two other goals: achieving the efficient resolution of mass torts and responding to the perceived inability of juries to understand and apply complex scientific analyses. The Ninth Circuit's *Daubert* opinion is an example of this response. It is not surprising that the judiciary has reached for whatever tools are readily at hand in an attempt to deal with these problems. This Article has argued that restrictive admissibility rulings are not the best way to achieve these objectives. Restrictive admissibility rulings do resolve the problems created by the repeated litigation of the same factual question. They do so, however, at the cost of confusing the issues of sufficiency and admissibility and of casting a shadow of uncertainty over whether similar rules may be applied in other circumstances. To the degree courts do require more efficient ways of resolving individual cases in mature mass torts, they should develop sufficiency rules specific to this need.

The problem of jury comprehension of complex scientific arguments presents a more complex issue. Sometimes parties do attempt to introduce testimony so lacking in validity that exclusion is appropriate because the testimony threatens to cloud the issue and confuse the jury. Even with respect to Bendectin litigation, however, an area that some have pointed to as an example of "junk science,"²⁷⁹ it is difficult to justify the exclusion of the plaintiff's entire case on scientific validity grounds. More importantly, marginal science is not the primary source of jury difficulties with complex scientific arguments. The heart of that problem lies not in the arguments of expert witnesses but rather in the structures and processes of adversarial adjudication that systematically disadvantage the cultural values of science. It is there that we should seek a remedy.

279. HUBER, GALILEO'S REVENGE, *supra* note 13, ch. 7. *But see* Kenneth J. Chesebro, *Galileo's Retort: Peter Huber's Junk Scholarship*, 42 AM. U. L. REV. 1637 (1993) (criticizing Huber's conclusions).

