

1-1-1992

COMMENT: *DeLuca v. Merrell Dow Pharmaceuticals, Inc.*: Statistical Significance and the Novel Scientific Technique

Constantine Kokkoris

Follow this and additional works at: <https://brooklynworks.brooklaw.edu/blr>

Recommended Citation

Constantine Kokkoris, COMMENT: *DeLuca v. Merrell Dow Pharmaceuticals, Inc.*: Statistical Significance and the Novel Scientific Technique, 58 Brook. L. Rev. 219 (1992).

Available at: <https://brooklynworks.brooklaw.edu/blr/vol58/iss1/9>

This Note is brought to you for free and open access by the Law Journals at BrooklynWorks. It has been accepted for inclusion in Brooklyn Law Review by an authorized editor of BrooklynWorks.

COMMENT

*DeLuca v. MERRELL DOW PHARMACEUTICALS, INC.:** STATISTICAL SIGNIFICANCE AND THE NOVEL SCIENTIFIC TECHNIQUE

INTRODUCTION

Amy DeLuca was born with a deformed lower left leg: her tibia was bowed, her fibula and three toes were missing, her leg was too short and her right foot was missing a toe.¹ Amy's mother was prescribed Bendectin to alleviate morning sickness during her pregnancy. Alleging that the drug caused Amy's birth defects, the DeLucas brought suit against the manufacturer.² They did so despite the absence of any conclusive medical evidence that the drug is a teratogen (an agent that causes birth defects).³ In fact, no published study has concluded that Bendectin is related to the type of birth defects suffered by Amy DeLuca.⁴

Bendectin was approved for sale in 1956 by the Food and Drug Administration ("FDA").⁵ In 1980 the FDA re-examined the drug's safety in response to mounting public concern that Bendectin was a teratogen.⁶ The FDA's Advisory Committee on Fertility and Maternal Health found no association between Bendectin and birth defects, but urged that research continue.⁷ Although the FDA has never withdrawn its approval of Bendec-

* 911 F.2d 941 (3d Cir. 1990).

¹ *DeLuca*, 911 F.2d at 943.

² *Id.*

³ *Id.* From more than 35 epidemiological studies, the Food and Drug Administration ("FDA") believes Bendectin is safe. Joanne Wojcik, *\$30 Million Punitive Award: Bendectin Maker Says Award Exceeds Texas Cap*, BUS. INS., Oct. 14, 1991, at 3.

⁴ *DeLuca*, 911 F.2d at 943.

⁵ *Id.*

⁶ *Id.*

⁷ *Id.*

tin, Merrell Dow Pharmaceuticals, the manufacturer, has withdrawn the drug from the United States market.⁸ The flood of Bendectin litigation across the nation, the increasing legal and insurance costs and the decrease in use of the drug because of adverse publicity, has caused Merrell Dow to cease production.⁹

To satisfy their burden of proving causation, the DeLucas offered the testimony of Dr. Alan Done, an expert in pediatric pharmacology.¹⁰ Dr. Done's opinion that Bendectin is a teratogen was based on four different types of evidence.¹¹ He conceded, however, that the only type on which a finding of causation in humans can be reliably based is epidemiological

⁸ *Id.* Richardson-Merrell Co. originally manufactured the drug. Richardson-Merrell later became a subsidiary of Dow Chemical Co. Ed Bruske, *Judge Overturns \$750,000 Award in Bendectin Case*, WASH. POST, Sept. 7, 1983, at D1, D4.

⁹ *DeLuca*, 911 F.2d at 943. After the first two Bendectin cases, that Merrell Dow won, the company offered to settle with all present and future claimants by creating a \$120 million fund. Over 700 claimants agreed to settle, but the Sixth Circuit overturned a class certification by the United States District Court for the Southern District of Ohio. The certification was objected to by some of the plaintiffs' attorneys. When it appeared that there was no way to get a certified agreement that would not give plaintiffs the right to opt out, Merrell-Dow withdrew the offer. *In re Bendectin Prods. Liab. Litig.*, 749 F.2d 300 (6th Cir. 1984). See also Philip E. Ross, *Drug Did Not Cause Birth Defects, Court Says*, N.Y. TIMES, Sept. 1, 1988, at A19.

In the first trial the jury awarded the plaintiff \$20,000 for medical expenses even though it found no causation. Merrell-Dow won on retrial. In the second trial the jury found for the plaintiff and awarded \$750,000, but was reversed by the court on a motion for judgment notwithstanding the verdict. Nonetheless, plaintiffs began bringing suits against Merrell-Dow in federal and state courts across the country. See *The Cause and Defect of Orangemail*, N.Y. TIMES, March 24, 1985, at A22 (editorial).

"As federal dockets swelled in the early 1980's with Bendectin cases, the Judicial Panel on Multi-District Litigation transferred over 600 of these cases to the Southern District of Ohio for pre-trial discovery, where they were consolidated with 557 cases filed within that district. See *In re Richardson-Merrell, Inc. Bendectin Prods. Liab. Litig.*, 624 F. Supp. 1212 (S.D. Ohio 1985), *aff'd in relevant part*, 857 F.2d 290 (6th Cir. 1988), *cert. denied*, 488 U.S. 1006 (1989)." *DeLuca*, 911 F.2d at 949.

¹⁰ *Id.* at 943. Dr. Done has testified in many Bendectin cases.

¹¹ *Id.* at 949. Dr. Done based his opinion on: (1) structure-activity analysis; (2) *in vitro* animal studies; (3) *in vivo* animal studies; and (4) his interpretation of epidemiological studies. Structure-activity analysis involves comparing the chemical structure of the drug to the structure of other known teratogens. He found that one of the components of Bendectin, an antihistamine, is a known teratogen in some animals and suspected of being a human teratogen. *In vitro* studies are test tube studies that use animal embryos to test the effect of the drug. *In vivo* studies test the drug on live animals. Although Dr. Done concluded that the three types of studies collectively suggested teratogenicity, he conceded that he could not infer Bendectin's human teratogenicity from any of these types of studies individually. *Oxendine v. Merrell Dow Pharmaceuticals, Inc.*, 506 A.2d 1100, 1104-07 (D.C. 1986).

evidence.¹² The epidemiological studies that Dr. Done based his opinion fell short of the level of statistical significance required by epidemiologists to establish causation.¹³ Merrell Dow argued that Dr. Done's testimony should therefore be excluded. Without any causation evidence, *DeLuca* would not survive summary judgment.¹⁴ The DeLucas argued that Dr. Done's testimony should be allowed because his epidemiological evidence was probative enough for purposes of legal fact finding, even if it failed to meet the more rigorous standards of the scientific community.¹⁵

Epidemiology is a branch of medicine that uses statistical information to track the incidence of disease throughout the population in an effort to determine causation.¹⁶ Epidemiological studies can provide evidence of causation where the biological cause of a disease is otherwise unknown.¹⁷ As with most birth defects, the cause of a child's limb reduction cannot be determined by clinical examination.¹⁸ Thus, in the Bendectin cases, courts have been deciding the issue of causation with purely statistical evidence.

The key issue before most Bendectin courts has been whether the expert testimony proffered by plaintiffs to establish causation is admissible.¹⁹ Most appellate courts addressing this issue have excluded the type of testimony offered by Dr. Done because it is not based on data "reasonably relied upon by experts in the particular field."²⁰ Implicit in these holdings is the

¹² *DeLuca*, 911 F.2d at 949 (citing *Oxendine*, 506 A.2d at 1108). Epidemiological studies observe the effect of the drug on the human population.

¹³ *Oxendine*, 506 A.2d at 1109.

¹⁴ *DeLuca*, 911 F.2d at 954.

¹⁵ *Id.* at 945.

¹⁶ Bert Black & David E. Lilienfeld, *Epidemiologic Proof in Toxic Tort Litigation*, 52 *FORDHAM L. REV.* 732, 750 (1984).

¹⁷ *DeLuca*, 911 F.2d at 945.

¹⁸ *Id.*

¹⁹ See, e.g., *DeLuca*, 911 F.2d at 949-952; *Ealy v. Richardson-Merrell, Inc.*, 897 F.2d 1159 (D.C. Cir. 1990); *Brock v. Merrell Dow*, 874 F.2d 307 (5th Cir. 1989); *Richardson v. Richardson-Merrell, Inc.*, 857 F.2d 823 (D.C. Cir. 1988); *Lynch v. Merrell-National Laboratories*, 830 F.2d 1190 (1st Cir. 1987); *Oxendine v. Merrell Dow*, 506 A.2d 1100 (D.C. 1986).

²⁰ *DeLuca*, 911 F.2d at 950. The *DeLuca* court stated:

We recognize that the district court's decision to exclude Dr. Done's proposed testimony was heavily influenced by the decisions of other courts that have grappled with the difficult question of whether expert testimony that Bendectin causes birth defects is admissible and/or sufficient to sustain a verdict.

assumption that, for an expert opinion on causation based on epidemiological studies to be admissible, the studies relied upon must show statistically significant results.²¹

In *DeLuca v. Merrell Dow Pharmaceuticals* the Third Circuit explicitly addressed the issue of whether statistically significant results are necessary for such testimony.²² The *DeLuca* court held that statistical significance should not automatically be set as the standard for admissibility simply because it is required by the scientific community.²³ The Third Circuit remanded the case to the district court for more extensive fact finding on the reliability of Dr. Done's methodology and its potential for confusing the jury. On remand the district court correctly found Dr. Done's methodology to be unreliable and, on that ground and others, excluded his testimony.²⁴

However, the Third Circuit could have provided a better rule of law regarding statistical significance and epidemiological evidence. Rather than suggest a generalized balancing test, the *DeLuca* court should have given specific content to its ruling by setting a clear standard as to when statistical significance should be required. The Third Circuit should have instructed the district court to admit only the non-significant epidemiological evidence if the district court determined that insufficient research existed on Bendectin's teratogenicity to make an informed decision. If insufficient research exists, plaintiffs should be allowed to proffer the best available current evidence on the causation issue.

This Comment will first set out a brief overview of epidemiological methodology. Next, it will analyze prior Bendectin case law, discussing the two approaches taken by courts to evaluate epidemiological evidence. The Comment will then discuss the *DeLuca* case and explain the Third Circuit's holding. It will argue that courts should not second-guess the epidemiological community's insistence on statistical significance for causation evidence so long as there is a body of scientific research that the

Id. at 949.

²¹ See *DeLuca*, 911 F.2d at 950; *Ealy*, 897 F.2d at 1162; *Richardson*, 857 F.2d at 829; *Lynch*, 830 F.2d at 1196-97. See also Michael Martin, *Issues About Experts in Toxic Tort Cases*, N.Y. L.J., Feb. 15, 1991, at 3.

²² Martin, *supra* note 21, at 7.

²³ *DeLuca*, 911 F.2d at 956-57.

²⁴ *DeLuca v. Merrell Dow Pharmaceuticals, Inc.*, 791 F. Supp. 1042 (D.N.J. 1992).

scientific community considers adequate to support a conclusion on causation. Finally, this Comment will propose a standard for the admissibility of epidemiological evidence that does not attain statistical significance.

I. BACKGROUND ON EPIDEMIOLOGY

"Epidemiology is the study of the distribution and determinants of disease in human populations."²⁵ Epidemiological studies seek "to observe the effect of exposure to a single factor upon the incidence of disease in two otherwise identical populations If the two [sample] groups are comparable, any difference in disease incidence can then be related either to the factor or to the sampling process, that is, to chance."²⁶

There are two general sources of error in statistical studies: non-random and random error.²⁷ Non-random error concerns the comparability of sample groups.²⁸ A study must sufficiently isolate the substance being tested to ensure that it is the only uniformly intervening variable distinguishing the exposed from the unexposed group. Only then can a finding of causation be reasonably inferred.²⁹

Non-random error is controlled by using proper experimental methods that can be verified by other epidemiologists when

²⁵ G. Marc Whitehead & Larry D. Espel, *Causation and Experts in Toxic Tort Litigation: How to Use and Defend Against Epidemiology, Toxicology and Biostatistics, in Preparation and Trial of A Toxic Tort Case* 1990, 387 PLI LITIG. & ADMIN. 169, 175 (1990), citing DR. ABRAHAM M. LILIENTFELD & DR. DAVID E. LILIENTFELD, *FOUNDATIONS OF EPIDEMIOLOGY* (2d ed. 1980).

²⁶ Black & Lilienfeld, *supra* note 16, at 755-57. Proper studies are designed to ensure that the sample groups selected are representative of the entire population. *Id.*

²⁷ CHARLES T. MCCORMICK, *MCCORMICK ON EVIDENCE*, §§ 208, 643, 644 (1984); Kenneth R. Kreiling, *Scientific Evidence: Toward Providing the Lay Trier With the Comprehensible and Reliable Evidence Necessary to Meet the Goals of the Rules of Evidence*, 32 ARIZ. L. REV. 915, 966-70 (1990).

²⁸ Black & Lilienfeld, *supra* note 16, at 755-59 (ideally the populations studied will be identical, except for the exposure); Whitehead & Espel, *supra* note 25, at 176; Otto Wong, *Using Epidemiology to Determine Causation in Disease, in Preparation and Trial of A Toxic Tort Case* 1990, 387 PLI LITIG. & ADMIN. 297, 299 (1990).

²⁹ See MCCORMICK, *supra* note 27, § 208, at 646 n.1 (correlation does not equal causation). This involves the validity of the study. Kreiling, *supra* note 27, at 973-75. It should also be noted that a proposed cause and effect relationship established by epidemiological studies should be biologically plausible (to the extent that current biological knowledge exists) to be fully accepted. Black & Lilienfeld, *supra* note 16, at 762; Wong, *supra* note 28, at 300.

studies are published in peer-reviewed scientific journals.³⁰ In this way, publication in a scientific journal may serve as a check on the validity of an epidemiological study.³¹ This Comment, however, will deal primarily with random error.

Random error involves the degree to which a disease occurs randomly throughout the population, independent of the substance being tested.³² It is likely that any two sample groups taken from a population will have differing incidence rates due strictly to variability of the population, i.e., chance.³³ In accounting for the difference between exposed and non-exposed groups, the epidemiologist must determine the probability that this difference is due strictly to chance and not to an effect of the exposure.³⁴ Although this possibility of random error cannot be completely eliminated, the risk can be controlled in statistical analysis by a process called "significance testing."³⁵

Next, to measure the strength of the association between the determinant and the disease, epidemiologists typically calculate the relative risk: "the ratio of the incidence rate of disease in the exposed group divided by that rate in the non-exposed control group."³⁶ Statistical tests are then conducted to determine the probability that the "difference in incidence rates results from sampling rather than from exposure."³⁷ These tests

³⁰ See Bert Black, *A Unified Theory of Scientific Evidence*, 56 *FORDHAM L. REV.* 595, 613-27 (1988) (acceptance as evidence of validity; publication as evidence of acceptance).

³¹ "The audience at which scientific publications are addressed is not passive; by its cheering or booing, its bouquets or brickbats, it actively controls the substance of the communications that it receives." John M. Ziman, *What is Science*, in *INTRODUCTORY READINGS IN THE PHILOSOPHY OF SCIENCE* 35, 40 (E.D. Klemke, R. Hollinger & A.D. Kline eds., 1980).

³² McCORMICK, *supra* note 27, § 208, at 644-46.

³³ *Id.* See also David H. Kaye, *Is Proof of Statistical Significance Relevant?* 61 *WASH. L. REV.* 1333, 1337-52 (1986).

³⁴ McCORMICK, *supra* note 27, § 208, at 644-46.

³⁵ *Id.* See also Kaye, *supra* note 33, at 1337-52.

³⁶ Black & Lilienfeld, *supra* note 16, at 757-58; Wong, *supra* note 28, at 298-301; Whitehead & Espel, *supra* note 25, at 176-79; *DeLuca*, 911 F.2d at 947. A relative risk of 1.0 would indicate no association; 2.0 would indicate that the incidence of disease doubled after exposure to the factor. Black & Lilienfeld, *supra* note 16, at 757-58.

Courts deciding causation based on epidemiological evidence alone have generally required a relative risk of over 2.0; this corresponds to the preponderance of evidence standard. A relative risk of 2.0 represents a 50% chance that anyone with the disease or birth defect got it from the substance in question. See *Id.* See also *Marder v. G.D. Searle & Co.*, 630 F. Supp. 1087, 1092 (D. Md. 1986).

³⁷ Black & Lilienfeld, *supra* note 16, at 757; McCORMICK, *supra* note 27, § 208, at

also determine the probability that the relative risk ratio reflects the true incidence ratio of the population.³⁸

A. Statistical Significance

Significance testing entails setting a "confidence interval" around the relative risk ratio within which the true disease incidence rate of the entire population can be expected to occur a certain percentage of the time.³⁹ The predominant standard used in scientific research is the 95% confidence interval; under this standard there is a 5% chance that the incidence rate will fall outside of the confidence interval and thus a 5% probability of error.⁴⁰ The larger the sample size of the study, the narrower the confidence interval.⁴¹ Narrow confidence intervals indicate that the estimate of relative risk is highly accurate; wide confidence intervals reveal more uncertainty.⁴²

644.

³⁸ See McCORMICK, *supra* note 27, § 203, at 645-46.

³⁹ Neil B. Cohen, *Confidence in Probability: Burdens of Persuasion in a World of Imperfect Knowledge*, 60 N.Y.U. L. REV. 385, 397-404 (1985); Kaye, *supra* note 33, at 1342-52; McCORMICK, *supra* note 27, § 203, at 645-46; KENNETH J. ROTHMAN, *MODERN EPIDEMIOLOGY* 119 (1986); DAVID W. BARNES & JOHN M. CONLEY, *STATISTICAL EVIDENCE IN LITIGATION* § 3.15 at 107 (1986).

⁴⁰ "The practice of using a certain standard levels of significance . . . can be traced to the influence of the eminent British statistician Sir R.A. Fisher . . ." As one contemporary statistician has remarked: "There you have it. Fisher thought 5% was about right, and who was there to disagree with the master?" DAVID S. MOORE, *STATISTICS: CONCEPTS AND CONTROVERSIES* 292 (1979).

⁴¹ Cohen, *supra* note 39, at 397-404. The larger the sample size of a study, the greater the "power" of the study is said to be.

⁴² To help clarify the concepts of confidence intervals, power and statistical significance, an example borrowed from a law review article on probability is useful:

Assume that we must determine the likelihood that the next marble drawn from an extremely large vat of black and white marbles will be white, given only the information that out of Y marbles randomly drawn from the vat in the past, X were white. If fifty prior drawings had produced thirty white marbles, our best guess for the probability of the next marble being white would be 0.6. If 100,000 marbles had been drawn, and had yielded 60,000 white marbles, our best estimate of the probability would again be 0.6. Nonetheless, these two probability assessments are quite different; although the estimates are identical, our confidence in their accuracy would differ dramatically due to the difference in the amount of information upon which they were based. In the first case, we would not be surprised if the actual probability of drawing a white marble turned out to be significantly different from 0.6, whereas in the second case a significant deviation from the 0.6 figure would be quite surprising.

This observation is not revolutionary. A professional statistician would not have described the two probabilities as identical. Rather, he or she would be

When the 95% confidence interval does not include a relative risk of 1.0, the result is "statistically significant" and a correlation between the factor and disease is established for scientific purposes.⁴³ One out of every twenty such studies would be expected to find a correlation when none actually exists.⁴⁴ The probability of error for falsely finding no correlation when one does exist is a different matter; although impossible to calculate precisely, it has been estimated at 50% when significance is tested with a 95% confidence level.⁴⁵ Thus roughly half of such studies would be expected to make the second type of error.⁴⁶

Strong evidence of correlation in a sound epidemiological study is generally considered circumstantial evidence of causation in the legal setting—causation may be properly inferred from it.⁴⁷ An example of a study yielding a statistically signifi-

more likely to say that, based on the information provided in the smaller sample, the probability that the next marble chosen will be white is ninety-five percent certain to be 0.6 plus or minus 0.14 (that is, between 0.46 and 0.74), whereas based on the information provided by the larger sample, the probability is ninety-five percent certain to be 0.6 plus or minus 0.003 (that is between 0.597 and 0.603). In making these statements, the statistician is describing "interval estimates," or "confidence intervals," in which it can be said with a specified level of confidence that the true value lies. Notably, the statistician could describe the probability of choosing a white marble in the second case as very likely to be greater than 0.5, but could not do so in the first case.

Cohen, *supra* note 39, at 398-99.

⁴³ Black & Lilienfeld, *supra* note 16, at 757 n.104; Whitehead & Espel, *supra* note 25, at 176-79; Cohen, *supra* note 39, at 398-404; Kaye, *supra* note 33, at 1348. A common misconception about confidence intervals is that they represent the probability that the results of a study are correct. A finding of statistical significance at the 95% confidence interval does not mean that it is 95% certain that exposure causes disease, with a 5% chance of error. Kaye, *supra* note 33, at 1347-49; McCormick, *supra* note 27, § 208, at 645; Cohen, *supra* note 39, at 397-99. Rather, significance testing represents the amount of confidence one can have in the relative risk ratio as an indicator of the probability of correlation. A study is more likely to yield statistically significant results as the sample groups become larger. See generally Cohen, *supra* note 39. This allows for narrower confidence intervals and hence greater accuracy in pinning down the true relative risk for the population. *Id.*

⁴⁴ Kaye, *supra* note 33, at 1342-45; Cohen, *supra* note 39, at 400-13.

⁴⁵ John M. Dawson, *Scientific Investigation of Fact—The Role of the Statistician*, 11 THE FORUM 896, 907 (1976).

⁴⁶ The first type of error (false positive) is referred to by statisticians as "type I" error. The second type (false negative) is referred to as "type II" error. See Cohen, *supra* note 39, at 411, 413.

⁴⁷ DeLuca, 911 F.2d at 945 ("Such studies have the potential, however, of generating circumstantial evidence of cause and effect . . ."). There is some disagreement among legal scholars on whether proof of increased risk of disease for a population should be

cant result is one that linked Bendectin to a birth defect of the stomach called pyloric stenosis. In this study, among 3,835 mothers exposed to Bendectin, thirteen had infants that developed pyloric stenosis. Among 9,511 women not exposed, thirteen also had infants that developed the malformation. This resulted in a relative risk ratio of 2.5. The 95% confidence interval stretched from 1.2 to 5.2.⁴⁸

Courts differ on whether epidemiological studies that fail to achieve statistical significance have probative value in the legal setting. Most courts have required statistical significance before admitting such proof to support a finding of causation.⁴⁹ The rationale of these decisions is that the scientific community requires statistical significance before accepting a finding of causation and that judges are in no position to contradict scientists when it comes to scientific research.⁵⁰

translated into proof of causation of the disease of a specific plaintiff. Cohen, *supra* note 39, at 385-86. For arguments that statistical information should be sufficient to support a verdict for the plaintiff in a civil case, see Cohen, *supra* note 39, at 419-22. See also Khristine L. Hall & Ellen K. Silbergeld, *Reappraising Epidemiology: A Response to Mr. Dore*, 7 HARV. ENVTL. L. REV. 441 (1983); Charles Nesson, *Agent Orange Meets the Blue Bus; Factfinding at the Frontier of Knowledge*, 66 B.U. L. REV. 521 (1986). But see Sargent v. Massachusetts Accident Co., 29 N.E.2d 825 (Mass. 1940); Michael Dore, *A Commentary on the Use of Epidemiological Evidence in Demonstrating Cause-in-Fact*, 7 HARV. ENVTL. L. REV. 429, 435-37 (1983); Lawrence Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329 (1971); Note, *Causation in Toxic Torts; Burdens of Proof, Standards of Persuasion, and Statistical Evidence*, 96 YALE L.J. 376 (1986).

⁴⁸ Pamela Aselton et al., *Boston Collaborative Drug Surveillance Program*, Boston University Medical Center, 1984 AM. J. EPIDEMIOL., 251, 251-56.

⁴⁹ Decisions that have rejected expert testimony on Bendectin's association with limb reduction because of lack of statistically significant findings include: *Ealy v. Richardson-Merrell, Inc.*, 897 F.2d 1159 (D.C. Cir. 1990); *Brock v. Merrell Dow*, 874 F.2d 307 (5th Cir. 1989); *Richardson v. Richardson-Merrell Inc.*, 857 F.2d 823 (D.C. Cir. 1988); *Lynch v. Merrell-National Lab.*, 830 F.2d 1190 (1st Cir. 1987).

Two decisions have accepted such epidemiological evidence as admissible: *Oxendine v. Merrell Dow Pharmaceuticals, Inc.*, 506 A.2d 1100 (D.C. 1986) (also accepting the evidence as sufficient to support a verdict), and *In re Richardson-Merrell, Inc. Bendectin Prods. Liab. Litig.*, 624 F. Supp. 1212 (S.D. Ohio 1985). However, the jury in the latter case found no causation and the court denied a post-trial motion for judgment notwithstanding the verdict by plaintiffs.

⁵⁰ Martin, *supra* note 21, at 4. See also Black, *supra* note 30, at 679-81; *In re Agent Orange Prods. Liab. Litig.*, 597 F. Supp. 740, 782-83 (E.D.N.Y. 1984); Jack B. Weinstein, *Improving Expert Testimony*, 20 U. RICH. L. REV. 473 (1986).

B. *The Epidemiological Evidence on Bendectin*

In the Bendectin cases, many courts have required epidemiological evidence to support a finding of causation.⁵¹ But the epidemiological evidence linking the drug to birth defects is far from strong. The "great weight of scientific opinion" rejects a link between Bendectin and increased risk of birth defects.⁵² Based on over thirty-five epidemiological studies conducted on Bendectin, the FDA has concluded that it was not a teratogen.⁵³ Of the thirty-five studies, four have suggested a link to birth defects at a statistically significant level, but only indicated an increased risk for pyloric stenosis (a stomach defect), heart defects and cleft palates.⁵⁴ No study has shown a statistically significant

⁵¹ *DeLuca*, 911 F.2d at 949-52; *Ealy*, 897 F.2d at 1163-64; *Brock*, 884 F.2d at 167; *Richardson*, 857 F.2d at 832; *Lynch*, 830 F.2d at 1194; *Oxendine*, 506 A.2d at 1108 (epidemiological evidence may be taken in combination with all other types in supporting an opinion on causation). See also *In re Agent Orange Prods. Liab. Litig.*, 611 F. Supp. 1223, 1231 (E.D.N.Y. 1985). But see *In re Bendectin Prods. Liab. Litig.*, 732 F. Supp. 744 (E.D. Mich. 1990), where the court held that epidemiological evidence was not necessary for an expert opinion on causation because:

[t]he 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.

Id. at 749 (citations omitted).

⁵² *DeLuca*, 911 F.2d at 945-46. See also Andrew Scolnick, *Key Witness Against Morning Sickness Drug Faces Scientific Fraud Charges*, 11 JAMA 1468, 1468 (1990).

⁵³ Joanne Wojcik, *supra* note 3.

⁵⁴ John de St. Jorre, *The Morning Sickness Drug Controversy*, N.Y. TIMES MAG., Oct. 12, 1980, at 113. Bruske, *supra* note 8, at D4. But as an FDA director conceded, it is almost impossible to get conclusive epidemiological evidence linking a drug with any specific birth defect.

Although the FDA approved Bendectin after a panel inquiry into the drug's teratogenicity in 1980, the panel admitted that the scope of the studies then available was insufficient to rule out the possibility that the drug was a "weak teratogen." Two studies linking the drug to heart defects and cleft palates left a "residual uncertainty" about the drug. The panel recommended: (1) continued research on the drug; (2) a package insert informing consumers of the current data; and (3) a warning to physicians that Bendectin should be used only for severe nausea and vomiting that cannot be treated without drugs. The FDA publicly announced the recommended warning and said that it was continuing to review animal and epidemiological studies. *Draft Guideline Patient Package Insert*, 45 Fed. Reg. 80740-01.

More evidence against Bendectin came late in 1982 and 1983 when two studies, one federally funded and one commissioned by the FDA, showed a statistically significant association between Bendectin and a type of stomach defect called pyloric stenosis. The

link between Bendectin and limb reduction. However, the number of children born with limb reduction defects is very small with respect to the number of birth defects in general; thus it is difficult to amass sufficient data on Bendectin and limb reduction alone.⁵⁵

In inferring causation from Bendectin studies in *DeLuca*, Dr. Done used a technique, advocated by Dr. Kenneth Rothman of the University of Massachusetts Medical School, that places diminished weight on significance testing.⁵⁶ Dr. Rothman contends that the .05 level is an arbitrary and conservative convention of epidemiology and that there may be a strong correlation between exposure and disease incidence without a finding of statistical significance.⁵⁷ He also contends that it is much more likely that the true incidence rate of the population falls toward the center of the confidence interval, rather than near the outer limits.⁵⁸ Thus, his approach focuses on the position of the whole confidence interval in relation to the relative risk, as opposed to focusing on the ends of the interval. Dr. Rothman advocates using several different confidence intervals, rather than just one.⁵⁹

Dr. Rothman also contends that his approach is more accurate than the traditional model in analyzing the results of more than one study.⁶⁰ Although each individual study may fail to at-

FDA was about to require Merrell Dow to disclose these findings in the official Bendectin labeling. However, on June 9, 1983, the company ceased production and distribution of the drug, claiming that the burden of marketing it had become too heavy because of the costs of defending about 300 lawsuits. At the same time, the company stated that there was no doubt as to the drug's safety. The FDA then requested that Merrell Dow send warning letters to doctors informing them of the studies for the benefit of those still using Bendectin. After the company refused, citing a third study with a larger database that showed no association with pyloric stenosis, the FDA alerted the nation's doctors itself. The agency included this third study in its bulletin, which emphasized that the available information could neither confirm nor disprove an association between Bendectin and the stomach defect. Morton Mintz, *FDA to Study 'Morning Sickness' Drug Link to Birth Defects*, WASH. POST, Aug. 23, 1980, at A7.

⁵⁵ According to an acting director of the FDA's Office of New Drug Evaluation, to establish a link between a drug and a specific birth defect, a study sample of many thousands of pregnant women is necessary (as opposed to several hundred for a "good study"). Michael De Courcy Hinds, *FDA May Warn on Drugs in Pregnancy*, N.Y. TIMES, Aug. 28, 1982, at A12.

⁵⁶ *DeLuca*, 911 F.2d at 946.

⁵⁷ ROTHMAN, *supra* note 39; Kenneth J. Rothman, *A Show of Confidence*, NEW ENG. J. MED., December 14, 1978, at 1362.

⁵⁸ *DeLuca*, 911 F.2d at 946.

⁵⁹ *Id.*

⁶⁰ *Id.*

tain significance, when viewed together, they may indicate a moderate or strong correlation between exposure and disease.⁶¹ This way, Dr. Rothman argues, the risk of type II error (false negative) is not as high as it is for significance testing at the .05 level.⁶² Also, by varying the size of the confidence interval, the researcher can adjust the study to better suit the type of decision being made and balance the risks of type I (false positive) and type II error.⁶³

Dr. Done's testimony relied on several studies, including one conducted by Merrell Dow.⁶⁴ He claimed to have found several errors in the method used to collect the data.⁶⁵ He also claimed to have corrected these errors and to have compiled data from all published studies on Bendectin and birth defects as well as data from several unpublished studies. Thus Dr. Done's database was broader than that of any other study. He also claimed to rely on the same data that Merrell Dow's experts (and the FDA) relied on in forming their opinions.⁶⁶ Using Dr. Rothman's approach to analyze these data, Dr. Done did not find significance at the .05 level, but found he could reject the null hypothesis at the .1 level (90% confidence interval).⁶⁷ He concluded that the "bulk of the available human epidemiological data . . . are indicative of Bendectin's human teratogenicity."⁶⁸

II. THE BENDECTIN CASE LAW

In considering the probative value of expert testimony proffered by Bendectin plaintiffs, most appellate courts have decided to look beneath the expert opinions and examine their underlying reasoning.⁶⁹ This is a departure from the traditional

⁶¹ *Id.*

⁶² *Id.*

⁶³ *Id.*

⁶⁴ *Oxendine v. Merrell Dow Pharmaceuticals, Inc.*, 506 A.2d 1100, 1107 (D.C. 1986).

⁶⁵ *Id.* That study, conducted by Drs. Bunde and Bowles—employees of Merrell Dow—used data from obstetricians on pairs of their female patients in Canada and the United States.

⁶⁶ *DeLuca*, 911 F.2d at 948.

⁶⁷ *Id.*

⁶⁸ *Id.* at 948-49 (quoting from appellant's brief). Dr. Done did not "quantify the increased risk for limb reduction defects," but he contended that the association of Bendectin with birth defects was strongest for, among other defects, limb reduction defects like Amy DeLuca's. *Id.* at 949.

⁶⁹ See *Ealy*, 897 F.2d 1159; *Brock*, 874 F.2d 307; *Richardson*, 857 F.2d 823; *Lynch*,

approach where courts would allow expert testimony without judicial scrutiny as long as the experts' qualifications as physicians were established and the experts expressed "reasonable medical certainty" in their opinions.⁷⁰

A. *The Traditional Approach: The Ferebee Standard*

An illustration of the traditional approach⁷¹ is *Oxendine v. Merrell Dow Pharmaceuticals Inc.*⁷² In *Oxendine* the District of Columbia Court of Appeals reversed a judgment notwithstanding the verdict ("j.n.o.v.") by the court below, holding that Dr. Done's epidemiological evidence taken together with his other causation evidence was sufficient for a jury's finding of causation.⁷³ The court noted that "Dr. Done's methodology was generally accepted in the field of teratology, and his qualifications as an expert have not been challenged,"⁷⁴ citing *Ferebee v. Chevron Chemical Co.*⁷⁵ *Ferebee* stands for the proposition that:

Judges, both trial and appellate, have no special competence to resolve the complex and refractory causal issues raised by the attempt to link low level exposure to toxic chemicals with human disease. 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 The case was thus a classic battle of the experts, a battle in

830 F.2d 1190.

⁷⁰ Black, *supra* note 30, at 659-62.

⁷¹ See also *Lanzilotti v. Merrell Dow Pharmaceuticals, Inc.*, 1986 WL 7832 (E.D. Pa. 1986).

⁷² 506 A.2d 1100 (D.C. 1986).

⁷³ As in *DeLuca*, the three other types of evidence presented were *in vivo* and *in vitro* studies, as well as structure activity analysis. Without epidemiological studies, these types of evidence are generally not considered to provide a sound basis for a finding of causation in the scientific community. *Oxendine*, 506 A.2d at 1110.

A similar result was reached in *In re Bendectin*, 732 F. Supp. 744 (E.D. Mich. 1990) (opinion written by Judge Carl Rubin, who also wrote the opinion in *In re Richardson-Merrell*, 624 F.Supp. 1212 (S.D. Ohio 1985), the consolidated class action), *Ramirez v. Richardson-Merrell*, No. 85-1504, 1986 WL 9724 (E.D. Pa. Sept. 4, 1986) and *Lanzilotti*, No. 82-0183 1986 WL 7832 (E.D. Pa. July 10, 1986). Those courts found a division in the scientific community as to whether epidemiological evidence is the only type that can support a conclusion of causation. Thus they held that an expert could reasonably rely on other types of evidence and the jury would have to decide on the credibility and weight of the testimony. *In re Bendectin*, 732 F. Supp. at 749; *Ramirez* at *3; *Lanzilotti* at *2.

⁷⁴ *Oxendine*, 506 A.2d at 1110.

⁷⁵ 736 F.2d 1529 (D.C. Cir. 1984).

which the jury must decide the victor.⁷⁶

Ferebee involved an agricultural worker who was exposed to the chemical paraquat on the job.⁷⁷ He alleged that as a result he subsequently contracted pulmonary fibrosis—a disease from which he later died.⁷⁸ His two treating physicians testified to causation on the basis of their own examination of him and medical studies they felt suggested that paraquat may cause chronic injury of the type afflicting *Ferebee*.⁷⁹ The defendant argued that while paraquat was known to be acutely toxic, there was no evidence linking it to chronic disease upon which the plaintiff's treating physicians could legitimately base an opinion of causation.⁸⁰ In appealing the trial court's denial of j.n.o.v., the defendant argued that "the jury was obligated to reject" *Ferebee*'s theory of causation.⁸¹ The court of appeals distinguished between an expert's methodology and an expert's conclusion,

⁷⁶ Thus, a cause-effect relationship need not be clearly established by animal or epidemiological studies before a doctor can testify that, in his opinion, such a relationship exists. As long as the basic methodology employed to reach such a conclusion is sound, such as use of tissue samples, standard tests, and patient examination, products liability law does not preclude recovery until a "statistically significant" number of people have been injured or until science has had the time and resources to complete sophisticated laboratory studies of the chemical. In a courtroom, the test for allowing a plaintiff to recover in a tort suit of this type is not scientific certainty but legal sufficiency; if reasonable jurors could conclude from the expert testimony that paraquat more likely than not caused *Ferebee*'s injury, the fact that another jury might reach the opposite conclusion or that science would require more evidence before conclusively considering the causation question resolved is irrelevant. That *Ferebee*'s case may have been the first of its exact type, or that his doctors may have been the first alert enough to recognize such a case, does not mean that the testimony of those doctors, who are concededly well qualified in their fields, should not have been admitted.

Id. at 1534-35. See also *Oxendine*, 506 A.2d at 1104, 1110.

⁷⁷ *Ferebee*, 736 F.2d at 1532.

⁷⁸ *Id.* at 1533. When *Ferebee* died before trial, his estate continued with a survival action and a wrongful death count was added on behalf of his minor children. *Id.* at 1532.

⁷⁹ *Id.* at 1533.

⁸⁰ *Id.* at 1535. However, the court noted that "[t]he dose-response relationship at low levels of exposure for admittedly toxic chemicals like paraquat is one of the most sharply contested questions currently being debated in the medical community . . . surely it would be rash for a court to declare as a matter of law that, below a certain threshold level of exposure, dermal absorption of paraquat has no detrimental effect." *Id.* at 1536 (citing James P. Leape, *Quantitative Risk Assessment in Regulation of Environmental Carcinogens*, 4 HARV. ENVTL. L. REV. 86, 100-03 (1980)).

⁸¹ *Ferebee*, 736 F.2d at 1535.

holding that only the methodology must be generally accepted in the scientific community.⁸² Thus the testimony of Ferebee's treating physicians was sufficient to support the jury verdict for the plaintiff.⁸³ This approach has been dubbed "passive acceptance" by one commentator.⁸⁴

The *Oxendine* court credited the testimony of Dr. Shanna Swan, an expert in biostatistics and epidemiology, offered by the plaintiff to rebut Merrell Dow's contention that statistical significance at the .05 level was necessary for a finding of causation in epidemiological studies.⁸⁵ The court also found that Dr. Done's use of *in vivo* and *in vitro* studies, chemical structure analysis and epidemiological studies in determining causation were "generally accepted in the field of teratology."⁸⁶ Finally, the court noted that Dr. Done's qualifications as an expert had not been challenged and stressed that "[a]lthough he was vigorously and exhaustively cross-examined by very able counsel, [Dr. Done] did not waiver from his opinion that Bendectin had caused appellant's birth defects."⁸⁷ Although the epidemiological studies that Dr. Done relied upon did not report significant results at

⁸² *Id.* The court stated that expert testimony could be controversial in its conclusions as long as it was based on "well-founded methodologies." *Id.* The methods the treating physicians used to diagnose Ferebee's illness were sound. However, their ability to determine the cause of Ferebee's pulmonary fibrosis by a medical examination was much more questionable. See Black, *supra* note 30, at 671-72.

⁸³ *Ferebee*, 736 F.2d at 1536.

⁸⁴ Black, *supra* note 30, at 671. This "passive acceptance" approach was also accepted by the Eleventh Circuit in *Wells v. Ortho Pharmaceutical Corp.*, 783 F.2d 741 (11th Cir. 1986). In *Wells* the plaintiff alleged that her use of defendant's spermicide during the first four weeks of her pregnancy caused birth defects in her child. Plaintiff's experts testified to causation on the basis of their expertise in the field, personal examinations of the child and medical and scientific studies. The epidemiological studies they relied on did not warrant any strong conclusions as to causation: they merely indicated the possibility. The defendant argued that stronger epidemiological proof was necessary for plaintiff's expert testimony to support a finding of causation by the trial court (the parties waived a jury). The Eleventh Circuit, citing *Ferebee*, held that, as a "battle of the experts," it was within the fact finder's province to determine the credibility of the witnesses and "decide the victor." *Id.* at 745 (citing *Ferebee*, 736 F.2d at 1535).

⁸⁵ *Oxendine*, 506 A.2d at 1109. The court held that since there was conflicting expert testimony on the necessity of .05 significance for a finding of causation, the issue "was properly left to the jury to resolve." *Id.* Accord *Lanzilotti v. Merrell Dow Pharmaceuticals*, 1986 WL 7832 (E.D. Pa. 1986) ("We believe that based solely on the testimony regarding the epidemiological studies, and the inferences that could be drawn therefrom, reasonable minds could disagree as to the teratogenicity of Bendectin.").

⁸⁶ *Oxendine*, 506 A.2d at 1110.

⁸⁷ *Id.* at 1108.

the .05 level, his testimony was still sufficient to support a verdict for the plaintiff.⁸⁸

B. *Criticism of the Ferebee Standard*

The *Ferebee* line of cases has been criticized by several commentators for allowing experts to testify to almost anything they wish, no matter how unscientific.⁸⁹ First, these critics fear that "an expert can be found to support almost any position"⁹⁰ and that, under the traditional approach, "courts cannot consistently and rationally resolve disputes about expert medical testimony."⁹¹ Second, commentators note that unsubstantiated expert testimony coupled with the presence of toxic tort victims, who make very sympathetic plaintiffs in the courtroom, may work great prejudice against defendant drug manufacturers.⁹² Others fear that the *Ferebee* standard has "the effect of turning over to doctors most of the decisions about legal sufficiency that

⁸⁸ *Id.* Three years after judgment was entered in *Oxendine*, Dr. Done was accused of misrepresenting his credentials during the *Oxendine* trial. The district court granted Merrell Dow's motion to vacate the judgment of *Oxendine I* and ordered a new trial. On appeal the court in *Oxendine II*, 563 A.2d 330 (D.C. 1989), found that Dr. Done had testified falsely that he was a faculty member of the medical school at Wayne State University as of May 1983; in fact, he had resigned before that date. But the court also held that information did not affect Dr. Done's qualification as an expert and that his standing in the medical community was still impressive. Therefore the misrepresentation was not material and the District of Columbia Court of Appeals once again reinstated the verdict.

⁸⁹ See Black, *supra* note 30, at 671-74, 677-81; Peter Huber, *Safety and the Second Best: The Hazards of Public Risk Management in the Courts*, 85 COLUM. L. REV. 277, 333 (1985); *Morning Sickness, Legal Miscarriage*, N.Y. TIMES, July 30, 1984, at A20 (editorial); Weinstein, *supra* note 50. See also *Richardson v. Richardson-Merrell, Inc.*, 649 F. Supp. 799 (D.D.C. 1986) (the district court criticized the *Oxendine* decision because it "judicially reopened an esoteric twenty-year-old controversy which is by now essentially settled within the scientific community.").

⁹⁰ Black, *supra* note 30, at 597-98 (quoting Huber, *supra* note 89 and Weinstein, *supra* note 50, at 482).

⁹¹ Black, *supra* note 30, at 669.

⁹² As the *Lynch* court noted: "The sight of a helpless mutilated youngster may evoke emotion along with the corresponding wish to make somebody pay for his or her plight. Judge Rubin has observed that the presence of handicapped youngsters could render a jury 'unable to arrive at an unbiased judgement.' *In re Richardson-Merrell*, 624 F. Supp. at 1224. With this very real possibility of runaway emotion overcoming judgment, the district court's firm rejection here of foundationless expert testimony was necessary . . ." *Lynch v. Merrell-National Labs.*, 830 F.2d at 1196. The court in *In re Richardson-Merrell*, 624 F. Supp. at 1222, excluded crippled minor plaintiffs from the courtroom against the plaintiffs' protest.

courts should make."⁹³

Finally, critics argue that the unpredictable jury verdicts which can arise where scientific testimony is admitted without regard for a uniform acceptance standard may also have a chilling effect on the pharmaceutical industry.⁹⁴ Critics point to manufacturers that have pulled useful drugs off the market for fear of tort liability and stopped research and development in the particular field.⁹⁵

C. *Active Review: Implicit And Explicit Adoption of the Frye Standard*

A number of courts faced with novel issues of causation in toxic tort cases have foregone the traditional approach and have scrutinized the reasoning of scientific experts before admitting their testimony.⁹⁶ Many of the Bendectin courts that have taken

⁹³ Black, *supra* note 30, at 670.

⁹⁴ See *Brock v. Merrell Dow Pharmaceuticals*, 874 F.2d at 307, 309-10 (5th Cir. 1989).

Under the traditional approach to scientific evidence, courts would not peer beneath the reasoning of medical experts to question their reasoning. Confronted, as we now are, with difficult medical questions, courts must critically evaluate the reasoning process by which the experts connect data to their conclusions in order for courts to consistently and rationally resolve the disputes before them. Moreover, in mass torts the same issue is often presented over and over to juries in different cases, and the juries often split both ways on the issue. The effect of this is to create a state of uncertainty among manufacturers contemplating the research and development of new, and potentially life-saving drugs. Appellate courts, if they take the lead in resolving those questions upon which juries will go both ways, can reduce some of the uncertainty which can tend to produce a sub-optimal amount of new drug development.

Id. See also Omnibus Trade and Competitiveness Act, 133 Cong. Rec. S10121-091 (daily ed. July 17, 1987)(report of Sen. McConnell) ("Development of new drugs is diminishing as product liability cases grow").

⁹⁵ See Malcolm Gladwell, *U.S. Firms Abandoning Birth Control Industry in Wake of Lawsuits*, WASH. POST, May 1, 1988, at J1; *Product Liability and U.S. Competitiveness*, 134 Cong. Rec. E3003-01 (daily ed. Sept. 16, 1988)(statement of Rep. Roth) ("One third of the companies surveyed have cancelled introduction of new products because of liability concerns").

⁹⁶ See *In re "Agent Orange" Prods. Liab. Litig.*, 597 F. Supp. 740, 783 (E.D.N.Y. 1984) ("It simply is not sufficient to point to an individual and show that he was exposed to Agent Orange and had a cancer."). See also *In re "Agent Orange" Prods. Liab. Litig.*, 611 F. Supp. 1223, 1260-63 (E.D.N.Y. 1985); *Johnston v. United States*, 597 F. Supp. 374, 415 (D. Kan. 1984) ("[The experts'] conclusions are not supported by any fact other than that the instruments are coated with a radioactive paint and each plaintiff has cancer This court is disappointed with the apparent fact that these so-called experts can take such license from the witness stand; these witnesses say and conclude things which,

this "active review"⁹⁷ approach cite Judge Weinstein's opinions in the Agent Orange litigation.⁹⁸

The Agent Orange litigation involved a class action by Vietnam War veterans who claimed that their exposure to the herbicide during the war caused them injury and caused birth defects in their children.⁹⁹ Judge Weinstein approved a settlement between plaintiffs and the manufacturing chemical companies, in part, because he found that plaintiffs would not meet their burden on the causation issue.¹⁰⁰ In dismissing the cases of those plaintiffs who had opted out of the class action for failing to provide *prima facie* evidence of causation, Judge Weinstein noted the lack of statistically significant findings of causation in any Agent Orange epidemiological study and the defects in the methodology of plaintiffs' expert witnesses.¹⁰¹ Although the epidemiological studies done on Agent Orange and dioxin (the active ingredient) did not confirm the plaintiffs' allegations, neither did they refute them.¹⁰²

Judge Weinstein recognized that the evidence provided by the plaintiffs supported the need for further research, but found that the evidence, at best, left the causation issue open.¹⁰³ He also noted that a definitive study of Agent Orange was underway, but its results would not be available in time for the litigation.¹⁰⁴ "Courts cannot, unfortunately, wait indefinitely until all

in the Court's view, they would not dare report in a peer-reviewed format.").

⁹⁷ This approach was dubbed "active review" by Bert Black. See Black, *supra* note 30, at 674.

⁹⁸ *Agent Orange*, 597 F. Supp. at 782-95, *Agent Orange*, 611 F. Supp. at 1230-63; *Lilley v. Dow Chemical Co.*, 611 F. Supp. 1267 (E.D.N.Y. 1985).

⁹⁹ *Agent Orange*, 597 F. Supp. at 746-48.

¹⁰⁰ *Id.*

¹⁰¹ *Agent Orange*, 611 F. Supp. at 1232-34. Judge Weinstein relied primarily on the Ranch Hand study on Agent Orange, conducted by the U.S. Air Force. *Id.* at 1232, 1241. He found "[t]he fact that the federal government was a defendant in related Agent Orange cases does not suggest a motive for untrustworthiness by the independent government scientists who conducted the studies." *Id.* at 1241. But see Nesson, *supra*, note 47, at 538. (suggesting that the Ranch Hand study may have been biased and criticizing Judge Weinstein for ignoring this problem, instead of basing "his reliance on the Air Force study with a similar study by the Australian Government, which had a similar problem with its men in Vietnam.").

¹⁰² *Agent Orange*, 611 F. Supp. at 1232; *Agent Orange*, 597 F. Supp. at 795.

¹⁰³ *Agent Orange*, 597 F. Supp. at 795.

¹⁰⁴ *Id.* at 782; *Agent Orange*, 611 F. Supp. at 1232-33. This study was to be undertaken by the Centers for Disease Control ("CDC"). Judge Weinstein noted that "[a]ssuming that the study was started in December, 1983 as planned . . . if all goes well

scientists have completed their long term studies. They must decide on information now available."¹⁰⁵ The court distinguished the Agent Orange cases before it from *Ferebee*, claiming that Ferebee's treating physicians "were relatively certain of the cause of his disease and no epidemiological proof was necessary," because no such studies existed.¹⁰⁶ By contrast, in the Agent Orange litigation, "no competent particularistic evidence has been presented and the relevant epidemiologic evidence is negative."¹⁰⁷

The approach taken by Judge Weinstein involved an implicit adoption of the "general acceptance" test promulgated by the United States Court of Appeals for the District of Columbia Circuit in *Frye v. United States*.¹⁰⁸ In *Frye* the defendant, charged with murder, sought to introduce a systolic blood pressure deception test (a precursor of the polygraph test)¹⁰⁹ to help prove his innocence.¹¹⁰ The deception test was relatively new at that time. Because it was a novel scientific technique that was not "sufficiently established to have gained general acceptance in the particular field in which it belongs,"¹¹¹ the court held the results inadmissible.¹¹²

it will be available in mid-1989." *Agent Orange*, 597 F. Supp. at 782. But all did not go well, and the study was abandoned because of difficulties in determining exposure. Philip M. Boffey, *Lack of Military Data Halts Agent Orange Study*, N.Y. TIMES, Sept. 1, 1987, at A1.

¹⁰⁵ *Agent Orange*, 597 F. Supp. at 782.

¹⁰⁶ *Agent Orange*, 611 F. Supp. at 1261-62.

¹⁰⁷ *Id.*

¹⁰⁸ 293 F.2d 1013 (D.C. Cir. 1923). See *Agent Orange*, 611 F. Supp. at 1242, 1261.

¹⁰⁹ It was based on the theory that conscious deception or concealment of guilt and fear of detection during examination raise the systolic blood pressure. *Frye*, 293 F.2d at 1013.

¹¹⁰ *Id.*

¹¹¹ *Id.* at 1014.

¹¹² 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.

Id. Years after the conviction, another person confessed to the murder. McCORMICK, *supra* note 27, at 606 n.5.

It is unclear whether the "thing from which the deduction is made" in *Frye* is the actual testing device, the theory of deception and blood pressure on which it is based, or the conclusion reached by the test. This is one of the problems with the application of

The rationale behind the "general acceptance" test is that "[a] courtroom is not a research laboratory" and "should not be used as a testing ground for theories supported neither by prior control experiments nor by calculations with indicia of reliability."¹¹³ Like Judge Weinstein in the Agent Orange litigation, many of the Bendectin courts have also implicitly adopted the *Frye* test.¹¹⁴ Courts granting summary judgment to Merrell Dow have neither characterized Dr. Done's approach as a "novel scientific technique" nor cited *Frye*.¹¹⁵ Yet their rejection of Dr. Done's testimony for his reliance on studies that did not attain significance at the .05 level implicitly accepts the *Frye* "general acceptance" standard.¹¹⁶ These courts assumed that since most experts in the field of epidemiology would not rely on Dr. Done's analysis (or on studies not significant at .05), the testimony was, therefore, not properly admissible as evidence of causation.¹¹⁷ Courts excluding Dr. Done's testimony determined, in effect, that the scientific community—and not the court—is best qualified to establish a threshold in inferring causation from test results.

The District of Columbia circuit court espoused such a view in *Richardson v. Richardson-Merrell, Inc.*¹¹⁸ The *Richardson* court noted that no published study found a statistically significant correlation between Bendectin and the type of birth defects

the *Frye* test that have emerged over the years. For criticism of *Frye*, see MCCORMICK *supra* note 27, at 605-07; Black, *supra* note 30, at 629-31; Paul C. Giannelli, *The Admissibility of Novel Scientific Evidence: Frye v. United States, A Half Century Later*, 80 COLUM. L. REV. 1197, 1208-21 (1980); Marc S. Klein, *Expert Testimony in Pharmaceutical Product Liability Actions*, 45 FOOD DRUG COSM. L.J. 393, 421-22. Many courts hold that it also applies to scientific techniques for analyzing data as well as forensic evidence. See *Manko v. United States*, 636 F. Supp. 1419, 1426 (W.D. Mo. 1986) (applying the *Frye* test to the use of the Mantel/Haenszel cohort analysis, "a recognized and generally accepted method of analyzing a stratified epidemiological data."); *Agent Orange*, 611 F. Supp. at 1261 (explaining the interpretation of the *Frye* rule in *Ferebee*: "The question is not whether the opinion itself is accepted in the relevant community, but instead whether the technique is. Inference from examination and testing, the court found, is clearly an accepted methodology.").

¹¹³ Klein, *supra* note 112, at 421.

¹¹⁴ Martin, *supra* note 21, at 4.

¹¹⁵ *Id.*

¹¹⁶ *Id.*

¹¹⁷ See *supra* note 20 and accompanying text.

¹¹⁸ 857 F.2d 823 (D.C. Cir. 1988). "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." *Id.* at 829.

at issue in the case. The court also mentioned Dr. Done's concession that significance at the .05 level was necessary to make a valid scientific conclusion as to causation. Moreover, he conceded that the studies he rejected and reanalyzed were published in peer-reviewed scientific journals, while his own recalculations were not.¹¹⁹ The court then distinguished *Richardson* from its previous decision in *Ferebee*, stressing that the *Ferebee* standard applied only "when the causation issue is novel and 'stands at the frontier of current medical and epidemiological inquiry'."¹²⁰ According to the *Richardson* court, Bendectin had been extensively studied and thus the causation issue was far from the "frontier" of epidemiological inquiry.¹²¹ The court then upheld the district court's grant of j.n.o.v. in Merrell Dow's favor under Federal Rule of Evidence 702.¹²²

The District of Columbia Circuit reaffirmed its position two years later in *Ealy v. Richardson-Merrell, Inc.*¹²³ The *Ealy* court cited *Richardson* extensively, again distinguishing *Ferebee* on the grounds that an overwhelming body of Bendectin research existed and that every study failed to establish causation.¹²⁴ The

¹¹⁹ *Id.* at 829-31.

¹²⁰ *Id.* at 832 (quoting *Ferebee*, 736 F.2d at 1534).

¹²¹ *Id.* "Uniquely to this case, the law now has the benefit of twenty years of scientific study, and the published results must be given their just due." *Id.* "When such [epidemiological] studies are available and relevant, and particularly when they are numerous and span a significant period of time, they assume a very important role in determinations of questions of causation." *Id.* at 830. The court discounted the relevancy of the three other types of evidence Dr. Done based his opinion on by noting they cannot prove causation in humans "in the face of the overwhelming body of contradictory epidemiological evidence." *Id.*

In *Turpin v. Merrell Dow Pharmaceuticals, Inc.*, 959 F.2d 1349 (6th Cir. 1992), the court took a different view, concluding that the 35 epidemiological studies it reviewed did not provide conclusive evidence to either support or refute a connection between Bendectin and birth defects. The court felt that these studies lacked sufficient power due to their small sample sizes. It granted summary judgment in favor of Merrell Dow, citing the plaintiff's failure to produce conclusive evidence of causation.

¹²² *Richardson*, 857 F.2d at 824.

¹²³ 897 F.2d 1159 (D.C. Cir. 1990).

¹²⁴ The court also quoted from *In re Agent Orange* stating that: "When the expert's . . . testimony lie[s] at the periphery of what the scientific community considers acceptable, special care should be exercised in evaluating the reliability and probative worth of the proffered testimony under Rules 703 and 403." *Ealy*, 897 F.2d at 1162 (quoting *In re Agent Orange*, 611 F. Supp. at 1242). The *Ealy* court went on: "Unlike the circumstances of *Ferebee*, the body of published epidemiological opinions on the subject at hand is extensive, indeed massive, and all such opinions point to the same conclusion." *Id.* at 1162. The court summed up *Richardson* and the case at hand as follows: "[U]nder Rule

court also provided an explanation for the discrepancy between the findings of the District of Columbia Court of Appeals in *Oxendine* and its own findings in *Richardson* and *Ealy*; *Oxendine* was decided in 1983 and between that time and the *Richardson* decision in 1986, two large epidemiological studies had been published, concluding that no relation existed between Bendectin and birth defects.¹²⁵ The *Ealy* court concluded by recognizing that future research may generate adequate epidemiological proof of causation but, until then, an expert opinion refuting the well-established scientific consensus would be inadmissible.¹²⁶

The First Circuit came to a similar conclusion in *Lynch v. Merrell National Laboratories*.¹²⁷ That court noted that studies on Bendectin's teratogenicity began in 1963 and that in over twenty years of research, "[n]o correlation, much less a causal connection, has been demonstrated between the use of Bendectin and limb reduction."¹²⁸ The *Lynch* court urged caution in allowing expert testimony on causation in a field like birth defects and stated that only a new (statistically significant) study challenging the consensus would be admissible evidence.¹²⁹

In *Brock v. Merrell Dow Pharmaceuticals*¹³⁰ the Fifth Circuit did not hold the testimony of plaintiff's expert (Dr. Glasser) inadmissible, but found it insufficient to support a jury verdict for the plaintiff.¹³¹ The court advocated active review for testimony on the causation issue. The *Brock* court noted that such testimony fosters unpredictable outcomes in jury trials, which, in turn, has a chilling effect on research and development of potentially helpful new drugs.¹³² Like the *Richardson* court, the *Brock* court noted that the expert's conclusions had not been

703, an opinion refuting this scientific consensus is inadmissible for lack of an adequate foundation, in the absence of other substantial probative evidence on which to base this opinion. It is this uncontroversial rule that is the ratio decidendi of *Richardson* and this case." *Id.* at 1162.

¹²⁵ *Id.*

¹²⁶ *Ealy*, 897 F.2d at 1163-64.

¹²⁷ 830 F.2d 1190 (1st Cir. 1987).

¹²⁸ *Id.* at 1194.

¹²⁹ *Id.*

¹³⁰ 874 F.2d 307 (5th Cir. 1989).

¹³¹ *Id.* Although the *Brock* court reversed the verdict on a sufficiency basis, "the underlying issue was really one of the admissibility of the experts' testimony." Martin, *supra* note 21, at 4.

¹³² *Id.* at 310. See *supra* note 94.

subject to peer review.¹³³ Although the court explicitly avoided holding that studies must be published to be admissible, the court, quoting *Richardson*, stated that "the examination of a scientific study by a cadre of lawyers is not the same as its examination by others trained in the field of science or medicine."¹³⁴

More recently, two circuits have explicitly adopted the *Frye* rule in passing on the admissibility of expert causation testimony. In *Daubert v. Merrell Dow Pharmaceuticals, Inc.*¹³⁵ the Ninth Circuit applied the general acceptance test in a particularly rigid fashion. The *Daubert* court held that an expert's methodology must meet all the essential requirements imposed by the scientific community. Accordingly, the court excluded plaintiff's causation evidence because it was generated solely for use in litigation and was not published or subjected to the normal peer review process.¹³⁶

In *Christopherson v. Allied-Signal Corp.*¹³⁷ the Fifth Circuit, sitting *en banc*, affirmed the district court's exclusion of the plaintiff's expert testimony on the causation of colon cancer (small cell carcinoma) by exposure to nickel and cadmium fumes. Exposure to nickel and cadmium fumes had been linked to small cell carcinoma of the lungs and other organs by epidemiological studies, but not to small cell carcinoma of the colon. Citing *Frye*, the *Christopherson* court required traditional proof

¹³³ *Id.* at 313.

¹³⁴ *Id.* (quoting *Richardson*, 857 F.2d at 831).

¹³⁵ 951 F.2d 1128 (9th Cir. 1991).

¹³⁶ The court stated that "the best test of certainty we have is good science—the science of publication, replication and verification, the science of consensus and peer review." *Daubert*, 951 F.2d at 1131, citing PETER HUBER, GALILEO'S REVENGE: JUNK SCIENCE IN THE COURTHOUSE 228 (1991).

There is disagreement on whether the drafters of the Federal Rules of Evidence intended to abandon the *Frye* test. There is no mention of it in the Rules. This has been interpreted both ways: "Probably the general scientific acceptance approach has survived the enactment of the Federal Rules, and will continue to be applied in determining the relevancy of such proof under Rule 401." DAVID W. LOUISELL & CHRISTOPHER B. MUELLER, FEDERAL EVIDENCE § 105, at 818. Accord STEPHEN A. SALTZBURG & KENNETH R. REDDEN, FEDERAL RULES OF EVIDENCE MANUAL 452 (3d ed. 1982). But see JACK B. WEINSTEIN & MARGARET A. BERGER, WEINSTEIN ON EVIDENCE § 3, at 702-36 (1992) ("[T]he silence of the rule [702] and its drafters should be regarded as tantamount to an abandonment of the general acceptance standard."). Accord C. WRIGHT & K. GRAHAM, 22 FEDERAL PRACTICE AND PROCEDURE § 5168, at 86-90 (1978). Many federal courts have rejected the *Frye* test. WEINSTEIN & BERGER, *supra*, §§ 702-36 to 702-37 n.6; McCORMICK, *supra* note 27, at 606-07 n.23.

¹³⁷ 939 F.2d 1106 (5th Cir. 1991).

(i.e., epidemiological studies, animal testing or *in vitro* studies) of a link between the exposure and Christopherson's particular disease before allowing the plaintiff's expert to testify on causation.

In doing so, it reversed the Fifth Circuit's previous holding in *Christopherson v. Allied-Signal Corp.*¹³⁸ that limited the requirement of positive epidemiological studies to the Bendectin cases, but failed to require such proof for all toxic tort cases. Since there had been no studies testing the link between exposure to nickel and cadmium fumes and small cell carcinoma in particular, the plaintiff could not meet this requirement. Consequently the court denied his motion for summary judgment.

The *Christopherson* opinion mentioned in a footnote that an expert opinion which is controversial or unique may still be admissible under the general acceptance test if it is supported by meaningful information.¹³⁹ The court noted that "scientific truth has not so completely hardened as to prevent legitimate difference of true expert opinion in a particular concrete context."¹⁴⁰ However, the *Christopherson* court dismissed the plaintiff's expert's theory of causation as mere speculation. The dissent objected, arguing that knowledge, through traditional methods, on the relationship between nickel, cadmium and small cell carcinoma was scant and that "the causation issue [was] plainly not sufficiently investigated to warrant summary proceedings based on direct judicial precedent."¹⁴¹

D. Criticism of the General Acceptance Test

Supporters of "active review" argue that by stressing publication and peer review for scientific findings, courts more effectively guarantee the validity and accuracy of an expert's conclusions.¹⁴² They further argue that conclusions reported by an epidemiologist in a research context are much less susceptible to bias than those testified to in a litigation context.¹⁴³ Many com-

¹³⁸ *Id.* at 1135.

¹³⁹ *Id.* at 1106.

¹⁴⁰ *Id.* at 1134.

¹⁴¹ *Christopherson*, 939 F.2d at 1123 (citing *Ferebee*) (Reaveley, J., dissenting).

¹⁴² See Black, *supra* note 30, at 677-81; Kreiling, *supra* note 27, at 970.

¹⁴³ "An expert can be found to testify to the truth of almost any factual theory, no matter how frivolous . . . Juries and judges can be, and sometimes are, misled by the expert-for-hire." Weinstein, *supra* note 50, at 482. "Today practicing lawyers can locate

mentators agree that "the examination of a scientific study by a cadre of lawyers is not the same as its examination by others trained in the field of science or medicine."¹⁴⁴

The "general acceptance" test has come under attack, however, for adopting the conservative, rigorous standards of science that tend to impose a heavy burden of proof on plaintiffs.¹⁴⁵ In particular, this argument has been directed at the use of statistical significance as a threshold for proof in the legal setting.¹⁴⁶ Some argue that the legal standard of proof is different than the scientific standard.¹⁴⁷ At least one commentator has argued that the balance between type I (false positive) and type II (false negative) error in a significance test at the .05 level conflicts with the preponderance of evidence standard in civil cases.¹⁴⁸

Another argument leveled against the "general acceptance" test is that it should not be applied when there are insufficient

quickly and easily an expert witness to advocate nearly anything the lawyers desire." Michael A. Graham, *Expert Witness Testimony and the Federal Rule of Evidence: Insuring Adequate Assurance of Trustworthiness*, 1986 U. ILL. L. REV. 43, 45. However, research scientists may not be completely immune from charges of bias. Allegations that some scientists slant their experiments to obtain results that are publishable have led to recent congressional inquiry. See HOUSE COMM. ON GOV'T OPERATIONS, ARE SCIENTIFIC MISCONDUCT AND CONFLICTS OF INTEREST HAZARDOUS TO OUR HEALTH? H.R. Rep. No. 688, 101st Cong., 2d Sess. (1990).

¹⁴⁴ Richardson v. Richardson Merrell Inc., 857 F.2d 823, 831 n.55 (D.C. Cir. 1993) (quoting from *Perry v. United States*, 755 F.2d 888, 892 (11th Cir. 1985)). Several commentators have called for upholding scientific standards in the court. See Black, *supra* note 30; Huber, *supra* note 89; Weinstein, *supra* note 50.

¹⁴⁵ See McCORMICK, *supra* note 27, at 608 ("'General scientific acceptance' is a proper condition for taking judicial notice of scientific facts but not a criterion for the admissibility of scientific evidence.").

¹⁴⁶ See McCORMICK, *supra* note 27, at 648-49.

One problem with [the use of significance testing] in court is the tendency of some expert witnesses or judges to assume that because there is an arbitrary convention of insisting on [significance] of .05 or less before labelling scientific findings "statistically significant," this same number should be required before the factfinder may rely on the quantitative results It is not the statistic that directly states how large the observed underrepresentation is. It is merely one measure of the probative force of statistical evidence, and an incomplete measure at that.

Id. See also Black & Lilienfeld, *supra* note 16, at 736 n.8; Kaye, *supra* note 36; David H. Kaye, *Statistical Significance and the Burden of Persuasion*, 46 LAW & CONTEMP. PROBS. 13 (1983); Cohen, *supra* note 39, at 409-17.

¹⁴⁷ Cohen, *supra* note 39, at 412-17; Kaye, *supra* note 33, at 1345; Nesson, *supra* note 47, at 529-32; Black, *supra* note 30, at 677-79. See also Richardson v. Richardson-Merrell Inc., 857 F.2d 823, 832 (D.C. Cir. 1988).

¹⁴⁸ Kaye, *supra* note 146.

data to accept or reject the null hypothesis with any degree of certainty.¹⁴⁹ One commentator criticized Judge Weinstein's decision to dismiss the cases of the opt-out plaintiffs in the Agent Orange litigation: Judge Weinstein dismissed these cases for failure to provide proof of causation, yet acknowledged that the epidemiological evidence indicated the need for further research, not the absence of causation.¹⁵⁰

In its decisions in *Ferebee* and *Richardson* the District of Columbia Circuit also recognized the potential unfairness to plaintiffs when the causation issue is novel and the proof that exists is insufficient for scientific purposes. The *Ferebee* court held that "products liability law does not preclude recovery until a 'statistically significant' number of people have been injured or until science has had the time and resources to complete sophisticated laboratory studies of the chemical."¹⁵¹ The *Richardson* court distinguished *Ferebee* as applying only to cases in which the causation issue is novel and "at the frontier of epidemiological inquiry."¹⁵² The *Richardson* court therefore excluded the expert testimony offered, finding that Bendectin had been extensively studied and was neither novel nor at the frontier of epidemiological inquiry. The *Richardson* court also specified that published epidemiological studies assume an important role in the causation issue when they are "numerous and span a significant period of time."¹⁵³

III. THE *DeLuca* OPINION

A. *The District Court Decision*

The district court excluded Dr. Done's testimony under Federal Rule of Evidence 104(a), holding that it did not meet "the foundational requirements of Fed. R. of Evid. 703."¹⁵⁴ Fed-

¹⁴⁹ See Nesson, *supra* note 47, at 529-32.

¹⁵⁰ *Id.*

¹⁵¹ *Ferebee v. Chevron Chemical*, 736 F.2d 1529, 1536 (D.C. Cir. 1984).

¹⁵² *Richardson v. Richardson-Merrell, Inc.* 857 F.2d 823, 832 (D.C. Cir. 1988); see *supra* notes 120-21 and accompanying text.

¹⁵³ 857 F.2d at 830.

¹⁵⁴ *DeLuca v. Merrell Dow Pharmaceuticals*, 131 F.R.D. 71, 72, (D.N.J. 1989), 911 F.2d 941 (3d Cir. 1990). The court cited other Bendectin cases in which summary judgment or j.n.o.v. was granted in favor of Merrell Dow, two of which involved testimony by Dr. Done himself. *Id.* at 73.

Federal Rule of Evidence 104(a) covers questions of admissibility. It states:

eral Rule of Evidence 703 addresses the admissibility of data upon which testimony by experts is based.¹⁵⁵ The *DeLuca* court explained that Dr. Done had not come forward with any new evidence of Bendectin's teratogenicity. Moreover, the court noted Dr. Done had conceded that the scientific community generally agreed that Bendectin does not cause birth defects.¹⁵⁶ It also explained that the epidemiological studies Dr. Done used failed to link Bendectin to limb reduction and that the other types of studies he relied on (i.e., *in vivo*, *in vitro* and chemical structure analysis) were "merely screening tests with no conclusive weight."¹⁵⁷ The court also mentioned that Dr. Done was not an epidemiologist.¹⁵⁸ Thus the *DeLuca* court concluded that his testimony contravened Rule 703 because it was not based on evidence reasonably relied upon by experts in the particular field.¹⁵⁹ Having excluded the *DeLuca*'s sole causation evidence, the court granted summary judgment in favor of Merrell Dow.¹⁶⁰

Preliminary questions concerning the qualification of a person to be a witness, the existence of a privilege, or the admissibility of evidence shall be determined by the court, subject to the provisions of subdivision (b). In making its determination it is not bound by the rules of evidence except those with respect to privileges.

FED. R. EVID. 104(a).

¹⁵⁵ Federal Rule of Evidence 703 states:

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

The rule was designed to allow experts to base their opinions on data like epidemiological studies that are considered hearsay and thus are inadmissible at trial. Before Rule 703, experts could only base their opinions on facts of which they had personal knowledge and facts in the trial records. WEINSTEIN & BERGER *supra* note 136, §§ 703-5 to 6 (1992).

¹⁵⁶ *DeLuca*, 131 F.R.D. at 74.

¹⁵⁷ *Id.*

¹⁵⁸ *Id.*

¹⁵⁹ *Id.* According to the Third Circuit, Rule 703, in conjunction with Rule 104(a), requires the court to "make a factual inquiry . . . as to what data experts in the field find reliable," and "[t]he proper inquiry is not what the court deems reliable, but what experts in the relevant discipline deem it to be." *In re Japanese Electronic Products Antitrust Litigation*, 723 F.2d 238, 276 (3d Cir. 1983), *rev'd on other grounds*, 475 U.S. 574 (1986). See also WEINSTEIN & BERGER, *supra* note 136, §§ 703[03], 703-24—703-25.

¹⁶⁰ *DeLuca*, 131 F.R.D. at 74.

B. *The Third Circuit Decision*

The Third Circuit reversed the district court's holding on Rule 703.¹⁶¹ The court noted that although Dr. Done was not an epidemiologist, Merrell Dow had conceded that he was qualified to interpret epidemiological studies.¹⁶² The court then found that Dr. Done relied on the same data relied on by Merrell Dow's experts—Dr. Done merely came to a different conclusion.¹⁶³ The *DeLuca* court distinguished between the data experts rely upon and the inferences they draw from the data, holding that Rule 703 only applied to the former.¹⁶⁴ Moreover, the court held that an expert need not accept the conclusion of a study to use the underlying data as a basis for testimony.¹⁶⁵ Thus Dr. Done's testimony was based on data reasonably relied upon by experts in the field.¹⁶⁶

According to the Third Circuit, the inquiry into experts' qualifications and the methodology they use in drawing conclusions from data are properly governed by Federal Rule of Evidence 702.¹⁶⁷ Rule 702 deals with the admissibility of expert testimony in general.¹⁶⁸

¹⁶¹ *DeLuca v. Merrell Dow Pharmaceuticals*, 911 F.2d 941, 952-54 (3d Cir. 1990).

¹⁶² *Id.* at 952.

¹⁶³ *Id.*

¹⁶⁴ Rule 703 is satisfied once there is a showing that an expert's testimony is based on the type of data a reasonable expert in the field would use in rendering an opinion on the subject at issue; it does not address the reliability or general acceptance of an expert's methodology. When a statistician refers to a study as 'not statistically significant,' he is not making a statement about the reliability of the data used, rather he is making a statement about the propriety of drawing a particular inference from that data He is making a statement about the degree to which the relationship found in the data may be due to chance, but his decision to use a certain significance level as a check on the permissible inference to be drawn from the data is a methodological value judgment which is separate from the question of whether the data is of the type an expert would rely upon.

Id. at 953.

¹⁶⁵ "[T]he Federal Rules of Evidence contain no requirement that an expert's testimony be based upon reasoning subjected to peer-review and published in the professional literature." *Id.* In *In re Paoli R.R. Yard PCB Litig.*, 916 F.2d 829 (3d Cir. 1990), the Third Circuit reaffirmed this part of the holding.

¹⁶⁶ *DeLuca*, 911 F.2d at 954.

¹⁶⁷ *Id.* at 954-57.

¹⁶⁸ Federal Rule of Evidence 702 states: "If scientific, technical or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise." FED. R. EVID. 702.

After briefly discussing epidemiology and the decisions of other Bendectin courts on the admissibility and weight of Dr. Done's testimony, the Third Circuit explicitly rejected the general acceptance test and substituted its own formulation from *United States v. Downing*.¹⁶⁹ In doing so, it characterized Dr. Done's analysis as a "novel scientific technique."¹⁷⁰

In *Downing* the Third Circuit held that the trial court erred in excluding defendant's expert testimony on the reliability of eyewitness testimony by not considering its helpfulness under Rule 702.¹⁷¹ The *Downing* court set out a three-part test for ruling on the admissibility of novel scientific evidence. Under this test, the trial court must determine: "(1) the soundness and reliability of the process or technique used in generating the evidence, (2) the possibility that admitting the evidence would overwhelm, confuse, or mislead the jury, and (3) the proffered connection between the scientific research or test result to be presented, and particular disputed factual issues in the case."¹⁷²

In *DeLuca* the Third Circuit remanded the case to the district court with instructions to consider the admissibility of Dr. Done's testimony in light of the *Downing* factors.¹⁷³ The court

¹⁶⁹ 753 F.2d 1224 (3d Cir. 1985). The *Downing* test uses the "relevancy" approach to admissibility of evidence, an approach advocated by McCormick, that has been termed "consistent with the underlying policies" of the Federal Rules of Evidence by Weinstein and Berger. See MCCORMICK, *supra* note 27, at 606-09; WEINSTEIN AND BERGER, *supra* note 136, at §§ 702-35 to 41. This approach is considered the rival to the *Frye* test. See Black, *supra* note 30, at 641-44; Kreiling, *supra* note 27, at 924.

¹⁷⁰ *DeLuca*, 911 F.2d at 954.

¹⁷¹ *U.S. v. Downing*, 753 F.2d 1224, 1232 (3d Cir. 1985). In *Downing* the defendant was charged with participating in a scheme to defraud numerous vendors. The participants attended trade shows, representing themselves as the Universal League of Clergy ("ULC"). They took orders from manufacturers' representatives and gave the representatives a list of supposed credit references that was actually a list of false banks and companies with the addresses of mail-drops from which ULC could collect the letters. After sending the manufacturers positive credit references, the manufacturers shipped goods to the ULC on credit that disposed of them without making payment.

The government's case against the defendant consisted almost entirely of 12 eyewitnesses who testified that the defendant represented himself as Reverend Claymore at the trade shows on behalf of ULC. The defendant denied representing himself as Reverend Claymore and asserted that Claymore was the real perpetrator. He sought to introduce the testimony of a psychologist on the unreliability of eyewitness testimony. The district court held that such expert testimony could never be admissible, but the Third Circuit reversed and remanded with instructions to consider the "helpfulness" of the testimony under Rule 702. *Id.*

¹⁷² *Downing*, 753 F.2d at 1237.

¹⁷³ *DeLuca*, 911 F.2d at 956-57.

made clear that the lack of statistical significance would be relevant in considering admissibility, but would not, in itself, require exclusion.¹⁷⁴ Additionally, even if the testimony was to be admitted, lack of statistical significance would be relevant in considering the sufficiency of the evidence to withstand summary judgment.¹⁷⁵ The *DeLuca* court also noted that the third prong of the *Downing* test was satisfied because epidemiological evidence is closely related to causation.¹⁷⁶ Thus the district court would have to consider only the reliability of the evidence and jury comprehension issues.¹⁷⁷

The Third Circuit then specified that in a motion for summary judgment the calculated relative risk would have to exceed 2.0 to satisfy the preponderance of the evidence standard.¹⁷⁸ The court also discussed the predominance of the .05 level of signifi-

¹⁷⁴ In considering the question of reliability on remand, the district court is permitted to identify relevant scientific communities and make determinations about the degree of acceptance of Dr. Done's methodology within those communities. Conversely, it may consider the extent to which members of these communities decline to give any weight to inferences not supported by .05 statistical significance. The district court should keep in mind, however, that the ultimate touchstone is helpfulness to the trier of fact, and with regard to reliability helpfulness turns on whether the expert's technique or principle [is] sufficiently reliable so that expert opinion with somewhat less assurance is not sufficiently reliable to be helpful in the context of civil litigation.

Id. (citations omitted).

¹⁷⁵ "[A]ssuming that New Jersey would apply the traditional 'more probable than not' burden of proof standard to the causation issue in this case, this admissible testimony would not alone bar summary judgment for Merrell Dow unless it would support a jury finding that Bendectin *more likely than not caused* the birth defects in *this particular case*" *Id.* at 958 (emphasis in original) (footnotes omitted). "Even if Dr. Done's statistical analysis is found to be admissible, its lack of statistical significance at the .05 level may appropriately play some role in deciding this subsequent issue." *Id.* at 959 n.24.

¹⁷⁶ *Id.* at 955.

¹⁷⁷ *Id.*

¹⁷⁸ *Id.* at 958-59 (quoting from *Manko v. United States*, 636 F. Supp. 1419, 1434 (W.D. Mo. 1986)). The *DeLuca* court explained that a relative risk of 2.0 indicates a doubling of disease incidence after exposure. This would indicate an attributable risk of 50%. Attributable risk is the probability that any person in the population born with birth defects acquired them because of exposure to the drug as opposed to other (background) causes. Only if the attributable risk is greater than 50% can it be said that it is more likely than not that the drug caused harm to any particular plaintiff. *Id.* See also Black & Lilienfeld, *supra* note 16, at 767-69.

Attributable risk is the percentage of people in the diseased population who can attribute their illness to the drug. The probability that any individual in this population was made ill by the drug (assuming exposure) is directly related to this percentage. See *id.* at 760-61, 767-69.

cance in scientific research.¹⁷⁹ The court cited Dr. Rothman's work,¹⁸⁰ noting his contention that "there is nothing magical or inherently important about .05 significance" and that "the data in a certain study may indicate a strong relationship between two variables but still not be 'statistically significant' and that the level of significance which should be required depends on the type of decision being made and the relative values placed on avoiding the two types of risk."¹⁸¹ The Third Circuit also explained Dr. Rothman's proposal to report the relative risk ratio and various confidence intervals (e.g., 90%, 95% and 99%), instead of reporting merely whether significance was attained.¹⁸²

After devoting a considerable portion of the opinion to the views of Drs. Done and Rothman, the Third Circuit instructed the district court to consider the degree to which the relevant scientific community accepts Dr. Done's methodology.¹⁸³ But the *DeLuca* court stressed that "helpfulness to the trier of fact . . . in reaching accurate results" was the central consideration and that even if scientists may require significance, it did not necessarily mean that "expert opinion with somewhat less assurance is not sufficiently reliable to be helpful in the context of civil litigation."¹⁸⁴

C. *The DeLuca Remand Decision*

On remand the district court held a five-day hearing in accordance with the principles set forth in the Third Circuit's

¹⁷⁹ *DeLuca*, 911 F.2d at 947. The court explained the concept of type I and type II error, and that when statistical significance is set at .05, the risk of type I error is 5%, while the risk of type II error may be as high as 50%. According to the court: "Type one error may be viewed here as the risk of concluding that Bendectin is a teratogen when it is not. Type two error is the risk of concluding that Bendectin is a teratogen when it is not. Type two error is the risk of concluding that Bendectin is not a teratogen, when it in fact is." *Id.* (citing John M. Dawson, *Investigation of Fact—The Role of the Statistician*, 11 THE FORUM 896, 907-08 (1976)). See *supra* notes 46-48 and accompanying text.

The court here was not completely accurate. Type I error is the risk of concluding Bendectin is a teratogen when the evidence would tend to show no correlation between the drug and birth defects, and vice-versa. See Cohen, *supra* note 39, at 411, 413.

¹⁸⁰ *DeLuca*, 911 F.2d at 946.

¹⁸¹ *Id.* at 947.

¹⁸² *Id.* at 948.

¹⁸³ *Id.* at 956.

¹⁸⁴ *Id.* at 956-57. Here, in a footnote, the court drew a distinction between what is helpful in a civil versus a criminal context but declined to pursue the topic any further. *Id.* at 957 n.20.

DeLuca and *Downing* decisions. It excluded Dr. Done's testimony under Federal Rules of Evidence 702 and 703.¹⁸⁵ The opinion contained an extensive critique of Dr. Done's expert testimony. Noting Dr. Done's failure both to consider the weight of existing epidemiological studies and to weigh more heavily the studies with larger sample sizes, the court found his methodology lacking indicia of reliability. The court also found that much of Dr. Done's calculations were based on erroneous data, that his methods were unexplained and unverifiable by other experts and that his presentation of the results to the jury was confusing and misleading.¹⁸⁶

IV. ANALYSIS: THE PROPOSED STANDARD

The Third Circuit was correct in identifying Dr. Done's analysis as a novel scientific technique. However, it should have made the District of Columbia Circuit's "frontier of epidemiological inquiry" formulation a part of its analysis.¹⁸⁷ A determination that a causation issue is at the frontier of epidemiological inquiry should be a prerequisite to admitting testimony based on studies that do not attain statistical significance. In making this determination, trial courts should take into account the power of the studies by referring to medical and epidemiological peer-reviewed literature.¹⁸⁸ They may enlist the aid of a qualified epidemiologist, as provided under Federal Rule of Evidence 706, to determine whether epidemiological studies are large enough to provide meaningful results.¹⁸⁹

The *Richardson* court recognized that the *Ferebee* standard was too lenient—plaintiffs who brought claims essentially untested in the medical arena could take advantage of the average juror's lack of scientific knowledge. The *Richardson* court remedied this weakness by making a valid distinction between issues at the "frontier of epidemiological inquiry" and those established in the scientific community.¹⁹⁰ But courts should allow

¹⁸⁵ See *supra* notes 155, 168 and accompanying text.

¹⁸⁶ *DeLuca v. Merrell Dow Pharmaceuticals Inc.*, 791 F. Supp. 1042 (D.N.J. 1992).

¹⁸⁷ *Richardson v. Richardson-Merrell, Inc.*, 857 F.2d 823, 832 (D.C. Cir. 1988); see *supra* notes 120-21, 152 and accompanying text.

¹⁸⁸ For a discussion of "power," see *supra* notes 41-42 and accompanying text.

¹⁸⁹ *DeLuca*, 911 F.2d at 956. Rule 706 provides for court appointed experts.

¹⁹⁰ *Richardson v. Richardson-Merrell, Inc.*, 857 F.2d 823, 832 (D.C. Cir. 1988); see *supra* notes 120-21, 152 and accompanying text.

deviation from accepted scientific practice only when there is a sound policy reason at stake. Only if courts decide that there is insufficient research on the issue should they admit testimony based on a lower standard than that accepted by the epidemiological community. This is a more logical way to approach the problem of novel causation evidence.

The literature on Bendectin and limb reduction defects is sparse. As a result of this low incidence rate of limb reduction defects, the sample sizes and the power of the epidemiological studies are less than ideal.¹⁹¹ However, considering the comprehensive inquiry by the FDA panel in 1980 and the over thirty-five published studies on Bendectin and birth defects to date, the causation issue here is well out of the "frontier" of epidemiological inquiry.¹⁹² The district court, on remand, was correct in stressing the value of the epidemiological studies with greater power and in excluding Dr. Done's testimony.¹⁹³ At this stage in Bendectin litigation, statistical significance should be required before allowing an expert to testify to causation based on epidemiological studies.¹⁹⁴

A. *Scientific Versus Legal Perspectives on Statistical Evidence*

For the epidemiologist, type I error (false positive) is a

¹⁹¹ Hinds, *supra* note 55, at 12.

¹⁹² See Wojcik, *supra* note 3, at 3. The causation issue for Bendectin is no longer novel. "Nearly 7 years after the drug's withdrawal, says Richard Leavitt, director of science information for the March of Dimes Birth Defects Foundation . . . there is still a general consensus among teratologists that Bendectin was one of the best studied drugs of all time for use in pregnancy and the great preponderance of evidence generally exonerates it from any harmful effect." Scolnick, *supra*, note 52, at 1468.

One commentator, who has reviewed all of the epidemiological evidence on Bendectin, has concluded that the sample sizes in many of the studies allowed for "relatively powerful analyses." Joseph Sanders, *The Bendectin Litigation: A Case Study in the Life Cycle of Mass Torts*, 43 HASTINGS L.J. 301, 345 (1992). He concluded that it was highly unlikely that these studies would miss a correlation between the drug and birth defects if one existed. *Id.*

¹⁹³ DeLuca v. Merrell Dow Pharmaceuticals Inc., 791 F. Supp. 1042 (D.N.J. 1992).

¹⁹⁴ The Third Circuit's purpose in applying the *Downing* standard in *DeLuca* was to help establish and define a rule for future application and not to change the district court's result in that case. Similarly, after the Third Circuit's remand in the *Downing* case, the district court again excluded the expert testimony, after applying the new test set forth by the Third Circuit. *United States v. Downing*, 609 F.Supp. 784 (E.D. Pa. 1985). This determination by the district court in *Downing* on the second time around was upheld on appeal. *United States v. Downing*, 780 F.2d 1017 (3d Cir. 1985).

greater evil than type II error (false negative). A positive result will be reported and relied upon by subsequent researchers.¹⁹⁵ Negative results often go unreported and, in any event, will not provide a basis for a scientific conclusion.¹⁹⁶ Epidemiologists, like other scientists, postpone a report of causation until there is a relatively high level of confidence in their findings.¹⁹⁷ This conservative approach is consistent with science's penchant for accuracy and is perfectly reasonable, considering that scientists have the option of reserving judgment on causation issues.¹⁹⁸ The research may be replicated subsequently and the study may be improved to increase its validity.¹⁹⁹ This is why epidemiology, by convention, has set the significance level at .05, at which the probability of a type I error is 5% while that for a type II error may be as high as 50%. Given the opportunity for repeated studies, larger sample groups and improved designs, researchers can more easily marshal sufficient evidence to support their hypotheses at a high level of significance.²⁰⁰

In the context of civil litigation, however, requiring statistical significance for a novel causation issue imposes too great a burden on plaintiffs.²⁰¹ While the scientific community has a much stronger aversion to type I error than type II error, as noted above, the legal community is equally concerned with both types of error in civil litigation.²⁰² As Justice Harlan explained in his influential concurring opinion in *In Re Winship*,²⁰³ the "comparative social disutility" of a wrongful verdict

¹⁹⁵ Nesson, *supra* note 47, at 529-32; Cohen, *supra* note 39, at 412.

¹⁹⁶ Nesson, *supra* note 47, at 529 n.26; Office of Science and Technology Policy, Chemical Carcinogens: A Review of the Science and Its Associated Principles, 50 Fed. Reg. 10371-442 (March 1985) ("A high-quality negative epidemiological study, while useful, cannot prove the absence of an association between chemical exposure and human cancer.").

¹⁹⁷ See Nesson, *supra* note 47, at 529-30.

¹⁹⁸ *Id.* Cohen, *supra* note 39, at 412.

¹⁹⁹ See Kreiling, *supra* note 27, at 965-71; Black, *supra* note 30, at 618 n.101.

²⁰⁰ As the sample size of a study increases, the confidence interval narrows, and thus it becomes easier to obtain statistically significant results. Cohen, *supra* note 39, at 397-401.

²⁰¹ Kaye, *supra* note 33, at 1345, 1354; Kaye, *supra* note 146, at 20.

²⁰² Kaye, *supra* note 146, at 19-20; Cohen, *supra* note 39, at 413-14.

²⁰³ 397 U.S. 358 (1970) (proof beyond a reasonable doubt required during adjudicatory stage when a minor is charged with an act that would constitute a crime if committed by an adult).

Because the standard of proof affects the comparative frequency of these two types of erroneous outcomes [false conviction and false acquittal], the choice of

for a plaintiff is no greater than that of a wrongful verdict for a defendant in a civil trial.²⁰⁴

The plaintiff has only one opportunity to present a case for causation and courts do not have the option of reserving judgment: they must decide the issue conclusively.²⁰⁵ The plaintiff will not be allowed to follow with subsequent studies.

Plaintiffs may not find themselves at a disadvantage when the causation issue upon which they offer evidence has been thoroughly researched over a vast period of time.²⁰⁶ More con-

the standard to be applied in a particular kind of litigation should, in a rational world, reflect an assessment of the comparative social disutility of each In a civil suit between two private parties for money damages, for example, we view it as no more serious in general for there to be an erroneous verdict in the defendant's favor than for there to be an erroneous verdict in the plaintiff's favor. A preponderance of the evidence standard therefore seems peculiarly appropriate for, as explained most sensibly, it simply requires the trier of fact "to believe that the existence of a fact is more probable than its nonexistence before [he] may find in favor of the party who has the burden to persuade the [judge] of the fact's existence."

Winship, 397 U.S. at 371-72 (quoting F. JAMES, *CIVIL PROCEDURE*, 250-1 (1965))(Harlan, J., concurring).

In another landmark case, a unanimous Supreme Court held that the preponderance of the evidence standard requires the litigants to "share the risk of error in roughly equal fashion." *Addington v. Texas*, 441 U.S. 418, 423 (1979).

²⁰⁴ Criminal trials involve a different allocation of "comparative social disutility." *Winship*, 397 U.S. at 372 (Harlan, J., concurring). The social disutility of a false conviction is much greater than that of a false acquittal. This is why the reasonable doubt standard is used for criminal trials as opposed to the preponderance of evidence standard. *Id.*

²⁰⁵ Nesson, *supra* note 47, at 529-30; Michael D. Green, *The Paradox of Statutes of Limitations in Toxic Substance Litigation*, 76 CAL. L. REV. 965, 995-96 (1988).

The mandate of our legal system is] to resolve conflicts, even seemingly intractable conflicts, from which we cannot retreat when parties involve the machinery of the legal system. In this sense, the jury represents neither the ideal arbitration of the scientific conflict, nor its permanent resolution, but simply the essential voice of the community in solving one problem fairly brought before it. Let the experts settle the larger dispute in due time; we have cases to resolve.

Christopherson v. Allied-Signal Corp., 939 F.2d 1106, 1128-29 (5th Cir. 1991)(Reavely, J., dissenting).

²⁰⁶ For example, the link between asbestos and cancer has been extensively studied, as "[k]nowledge of the potential health hazards of asbestos also dates back to ancient times." *In re Joint Eastern and Southern District Asbestos Litig.*, 129 B.R. 710, 735 (E.D.N.Y. 1991). For a discussion of asbestos, see Green, *supra* note 205, at 994 ("The history of asbestos litigation demonstrates that as time passes, more and better evidence is uncovered about the industry's and particular defendant's knowledge concerning the dangers of asbestos products to the variety of circumstances in which people were exposed."). By contrast, little is known about the risk posed to humans by dioxin. See Fred

clusive results can be inferred from a vast literature of scientific studies. But asking plaintiffs to prove significance at the .05 level is unduly burdensome when the causation issue is new, the malady in question is rare, or there is otherwise insufficient data on the issue. As previously noted, epidemiological studies with small sample sizes tend to yield wide confidence intervals, making it more difficult to attain statistically significant results.²⁰⁷

Epidemiological evidence may have strong probative value yet fail to attain statistical significance.²⁰⁸ It is quite possible that a drug currently only suspected to be a teratogen may be proved harmful at a statistically significant level in the future.²⁰⁹ To preclude plaintiffs from presenting causation evidence on a drug may rob plaintiffs of a rightful verdict simply because the strong showing of proof required by epidemiology was not available at trial. Such an outcome may undermine public faith in the ability of the legal system to adjudicate toxic tort cases fairly.²¹⁰

The prospect of insufficient evidence of causation is not implausible, even after there has been a considerable amount of time to study a drug. As noted, there are special difficulties involved in trying to link a drug to specific birth defects or other rare maladies because of the very low incidence rates involved.²¹¹

H. Tschirley, *Dioxin*, Sci. Am., Feb. 1986, at 29.

²⁰⁷ See *supra* notes 39-50 and accompanying text.

²⁰⁸ *DeLuca*, 911 F.2d at 948 ("it is 'much more likely that the [true] parameter [i.e. the true relationship between the studied variables] is located centrally within an interval than it is that the parameter is located near the limits of the interval'").

²⁰⁹ Nesson, *supra* note 47, at 537 (on the prospect that Agent Orange may be proved toxic in the future); Black & Lilienfeld, *supra* note 16, at 780; Green, *supra* note 205, at 975-76 (epidemiological studies are large, expensive undertakings and are often beyond the means of toxic tort plaintiffs). For the particular difficulty in determining drugs to be teratogens, see Hinds, *supra*, note 55, at 12, where it is compared to "searching for a needle in a haystack." The difficulty is in linking a drug to a specific type of birth defect. According to then acting director of the FDA's Office of New Drug Evaluation, "[t]heoretically, a good study of a drug to be used by pregnant women might involve a sample of several hundred pregnant women . . . but if you wanted to find out if the drug was associated with a specific defect you would need a study population of many thousands of pregnant women, and even if a drug was associated with the rare defect, you might not find it in this sample." *Id.*

²¹⁰ Nesson, *supra* note 47, at 537; Green, *supra* note 205, at 969, 1010.

²¹¹ According to an acting director of the FDA's Office of New Drug Evaluation, premarket studies cannot insure that drugs pose no risk to fetuses. Hinds, *supra* note 190, at 12. This problem is not limited to studies of birth defects; it applies to all rare defects/diseases.

The rarity of small-cell colon cancer creates virtually insurmountable obstacles

In addition, for obvious ethical reasons, suspected toxins or teratogens will not be administered on humans specifically for testing purposes.²¹² Therefore, in studying them, scientists often cannot conduct the preferred prospective studies, but must rely on retrospective studies, studying those who have already been exposed. These retrospective studies provide less certainty in determining causation.²¹³ Accordingly, it takes more time to amass sufficient proof of causation through epidemiological evidence when a drug is a suspected toxin.²¹⁴ Furthermore, sound epidemiological studies are often large, difficult and expensive under-

to statistically significant epidemiological studies. Accordingly, the majority of regimen creates virtually insurmountable obstacles to claimants suffering from rare or new diseases . . . The population has many times been exposed to unknown health hazards—with consequent injuries such as adenocarcinoma, pelvic inflammatory disease, toxic shock syndrome, and Guillian-Barre Syndrome—but a rigid alliance between law and epidemiology conspired to prevent recovery until a “statistically significant” number of deaths and injuries occurred. [citing Wendy E. Wagner, *Trans-Science in Torts*, 96 YALE L.J. 428, 429 (1986)]. Since the plaintiff in any event bears the burden of persuasion, and since the “more probable than not” causation standard already forces victims to bear significant losses without recompense, those cases adopting a less stringent view of scientific proof surely promotes the more enlightened and humane view of text law.

Christophersen, 939 F.2d 1106, 1128 n.19 (5th Cir. 1991)(Reavley, J., dissenting).

²¹² According to an acting director of the FDA's Office of New Drug Evaluation, “Ideally, we would like to have information on the drugs' effect on pregnant women, but it is an ethical question, and drug investigators cannot test experimental drugs on pregnant women . . .” As a result little is known about the effect most non-obstetric drugs have on human fetuses.” Hinds, *supra* note 190, at 12.

²¹³ See Green, *supra* note 205, at 995-96. A notable example is the Agent Orange research.

²¹⁴ Another problem with retrospective studies is the selective memory problem. In the case of Bendectin, mothers of children with birth defects are more likely to remember having taken the drug than other mothers. The potential for error of this type may harm the validity of study.

In *Brock* the court explained possible validity problems with the epidemiological studies, such as the “selective memory problem” and the possibility that another factor (smoking) may be causing the birth defects. *Brock v. Merrell Dow Co.*, 874 F.2d 307, 311-312 (5th Cir. 1989). The court then stated that “[f]ortunately, we do not have to resolve any of the above [problems with validity of the epidemiological studies], since the studies presented to us incorporate the possibility of these factors by use of a confidence interval.” *Id.* This statement is inaccurate. The confidence interval (statistical significance) is used to determine the probability that differences between the exposed and non-exposed groups are due to chance, assuming that the two groups are otherwise the same. The confidence interval cannot cure problems with the validity of a study. Black & Lilienfeld, *supra* note 16, at 756-56 n.104. However, there are other statistical tests available to “adjust the relative risk to account for the differences between [the exposed and non-exposed groups].” *Id.* at 756-57 n.103.

takings.²¹⁵ For example, the Agent Orange study mentioned by Judge Weinstein was never completed because of difficulties in determining human exposure to the herbicide.²¹⁶

Given these uncertainties, the best way to guarantee plaintiffs a fair trial on novel causation issues is to allow them to proffer the best available epidemiological evidence, regardless of its statistical significance at the .05 level. The priority of the legal system is to discover the truth of issues before it on the best available evidence. Courts should not hesitate to modify evidentiary rules so as to provide fact finders with more relevant evidence on which to base a decision.²¹⁷

B. *Using the Statute of Limitations to Improve Accuracy*

Ideally, there would seem to be a better approach for novel causation evidence than lowering the standard of admissibility for expert testimony.²¹⁸ The scientific uncertainty involved in a novel causation issue might be better dealt with by extending or eliminating the statute of limitations on the claim.²¹⁹ Allowing toxic tort plaintiffs to file their claims when causation has been firmly established by the scientific community provides a remedy for those who cannot immediately prove causation without compromising the quality of proof.²²⁰ Determinations of causation can then be based on more expansive and reliable information.²²¹ The delay that plaintiffs may suffer before recovering on

²¹⁵ Nesson, *supra* note 47, at 537-38. For what an epidemiological study involves, see *supra* notes 25-38.

²¹⁶ See *supra* note 104.

²¹⁷ As stated by Judge Fuld of the New York Court of Appeals in *Fleury v. Edwards*, 14 N.Y.2d 334, 341, 200 N.E. 2d 550, 557, 251 N.Y.S.2d 647, 654 (1964): "The common law of evidence is constantly being refashioned by the courts of this and other jurisdictions to meet the demands of modern litigation Absent some strong public policy or a clear act of pre-emption by the Legislature, rules of evidence should be fashioned to further, not frustrate, the truth-finding function of the courts in civil cases." The Federal Rules of Evidence also support a liberal view of admissibility. According to the Third Circuit, "there is a 'presumption of helpfulness' accorded expert testimony generally under Fed. R. Evid. 702." *United States v. Downing*, 753 F.2d 1224, 1241 (3d Cir. 1985) (quoting from *In re Japanese Antitrust Litig.*, 723 F.2d 238, 279 (3d Cir. 1983). The Federal Rules favor the admissibility of any relevant evidence. WEINSTEIN & BERGER, *supra* note 136, at § 702-44; McCORMICK, *supra* note 27, at 607.

²¹⁸ See generally Green, *supra* note 205.

²¹⁹ *Id.*

²²⁰ *Id.* at 970.

²²¹ *Id.* at 996-97, 1011-12.

their claims is a reasonable sacrifice for the relative certainty of the judgments that follow.²²²

However, there are several problems with such a formulation. First, the discovery rules currently used in most jurisdictions do not allow such a result.²²³ Almost all states that apply a discovery rule allow a statute to be tolled only until the plaintiff discovered, or should have discovered through due diligence, the cause of injury.²²⁴ These statutes do not allow for postponement until causation is established by the scientific community, but only until the plaintiff can identify the defendant as the source of the injury.²²⁵ The second problem in setting a discovery rule

²²² *Id.* at 968-69, 994-95.

²²³ *Id.* at 977-78, 982-84.

²²⁴ *Id.* at 983. For example, see New York's discovery rule, codified in N.Y. Civ. PRAC. L. & R. § 214-c (McKinney 1990). New York also enacted a discovery rule especially for Agent Orange, N.Y. Civ. PRAC L. & R. § 214-b. See Practice Commentaries at 627 (McKinney 1990). For examples of different formulations of the discovery rule in different states, see Green, *supra* note 205, at 978 n.63.

²²⁵ Green, *supra* note 205, at 983. The result in *Urland v. Merrell Dow Pharmaceuticals*, 822 F.2d 1268 (3d Cir. 1987), another Bendectin case decided by the Third Circuit, is illustrative of this problem. In that case the court upheld a jury verdict barring the plaintiff's claim on statute of limitations grounds. Although it used the discovery rule under the applicable state law (Pennsylvania) to toll the statute, the court held that the statute began to run when the plaintiff "possesse[s] the salient facts concerning the occurrence of his injury and *who or what* caused it . . ." *Id.* at 1275 (emphasis in original)(quoting *Staiano v. Johns-Manville Corp.*, 450 A.2d 681, 685 (Pa. 1982)).

The *Urlands'* knowledge of causation under this standard was deemed sufficient when they learned of the first Bendectin trial in Florida in 1980, *Mekdeci v. Merrell Nat'l Lab.*, 711 F.2d 1510 (11th Cir. 1983), which revived their suspicions that the drug had caused birth defects in their own child. Yet only three months later, the First Circuit in *Lynch v. Merrell National Lab.*, 830 F.2d 1190 (1st Cir. 1987), granted Merrell Dow's motion for summary judgment for the failure of plaintiffs to meet their burden of proof on causation. The results in these two cases seem inconsistent. See Green, *supra* note 205, at 983 n.82.

The Pennsylvania discovery rule was articulated by the Pennsylvania Superior Court as whether "the plaintiff knows, or reasonably should know: (1) that he has been injured, and (2) that his injury has been caused by another party's conduct." *Cathcart v. Keene Indus. Insulation*, 471 A.2d 493, 500 (Pa. Super. Ct. 1984). The Pennsylvania statute of limitations runs for two years. 42 PA. CONS. STAT. ANN. § 5524(2). A more recent Pennsylvania statute provides that the statute of limitations cannot run against a child during the period of minority (up to age 18). 42 PA. CONS. STAT. ANN. § 5533(b). However, *Urland* was brought to trial before the effective date of the statute and "the minor's tolling statute does not act to revive claims that had already been barred by the applicable statute of limitations prior to the effective date of the Act." *Urland*, 822 F.2d at 1276. New Jersey has such a minor tolling statute, N.J. REV. STAT. § 2A:14-21 (1987), that allowed the DeLucas to bring an action for their daughter, who was three years old at the time the action was brought, in New Jersey Superior Court. Brief for Appellants at 3-4, *DeLuca v. Merrell Dow Pharmaceuticals*, 911 F.2d 941 (3d Cir. 1990). The New Jersey

that tolls the statute of limitations until scientific proof is available is that disputes will arise over when the scientific evidence for causation is considered sufficient.²²⁶ In scientific circles there is seldom instant acceptance, but rather a long process of confirmation for a new theory.²²⁷ Litigation over the limitations issue might overshadow litigation on the substantive grounds, defeating the purpose of the statute.²²⁸ For this reason, one commentator has advocated a total elimination of statutes of limitations for toxic tort victims when science has not sufficiently passed on the connection between the injury and the alleged cause.²²⁹

C. Cost Shifting

Another problem with extending the statute of limitations to improve accuracy has to do with incentive. Epidemiological studies require large sample sizes to provide meaningful results.

statute of limitations for personal injury runs for two years. N.J. REV. STAT. § 2A:14-2 (1987).

Such inconsistent results under the discovery rule also occurred in the Agent Orange litigation. In *Fratlicelli v. Dow Chemical Co.*, 611 F. Supp. 1285 (E.D.N.Y. 1985), Judge Weinstein similarly held that under the Hawaii statute of limitations plaintiffs' actions accrued at the time they "knew of the act, the damage and the causal nexus." *Id.* at 1288. Plaintiffs' statements on a workers' compensation claim form that they knew of their disabilities resulting from exposure to the herbicide were enough to start the statute running. In the same opinion, Judge Weinstein rejected plaintiffs' causation argument for lack of admissible evidence. Thus the discovery rules applied in both *Urland* and *Fratlicelli* required a finding that plaintiffs had knowledge of the cause of their injuries before causation could be proved.

The Hawaii statute of limitations for injury to persons or property runs for two years. HAW. REV. STAT. § 657-7 (1991). The Supreme Court of Hawaii articulated that state's discovery rule as tolling the statute of limitations until the plaintiff discovers, or reasonably should have discovered, the "negligent act, the damage, and the causal connection between the former and the latter." *Yamaguchi v. Queen's Medical Center*, 648 P.2d 689, 693-94 (Haw. 1982).

The New York statute does take into account the "technical, scientific or medical knowledge" in determining causation for purposes of its discovery rule, but it tolls the statute for only a maximum of five years. N.Y. CIV. PRAC. L. & R. § 214-c (McKinney 1990). If, during this time, the plaintiff can "demonstrate that the state of medical or scientific knowledge was such that the causation of his injury could not have been identified" within the normal statutory period, the statute will begin to run for one year. Practice Commentaries, at 634-35 (McKinney 1990).

²²⁶ Green, *supra* note 205, at 983-84. See also N.Y. CIV. PRAC. L. & R. § 214-c, Practice Commentaries at 635 (McKinney 1990) (in reference to the New York discovery rule).

²²⁷ See Black, *supra* note 30, at 622-27; Kreiling, *supra* note 27, at 965-71.

²²⁸ Green, *supra* note 205, at 983-84.

²²⁹ *Id.*

Such studies are generally large and expensive undertakings, requiring institutional funding.²³⁰ Most toxic tort plaintiffs do not have the means or money to organize them and must rely on those studies conducted by others with a motive to incur such expense—usually the government or defendant manufacturers.²³¹

The government will generally undertake such an expense when there is strong suspicion that a widely used substance is toxic and likely to affect a considerable segment of the population.²³² A defendant drug manufacturer has little incentive to conduct epidemiological studies on a product that is allegedly toxic. For drug manufacturers, it is better to let the state of scientific knowledge on the product remain unclear than to provide potential plaintiffs with ammunition for a toxic tort suit.²³³

Allowing epidemiological studies into evidence that show an association between a drug and disease, but do not attain statistical significance because of small sample sizes, will provide defendant drug manufacturers with the incentive to commission studies large enough to provide a clearer picture on the causation issue.²³⁴ If the preliminary epidemiological evidence indicates that there may be a causal connection, corporate defendants will feel compelled to produce more reliable evidence negating this connection, rather than risk losing a verdict on the current, more speculative evidence.

Laying this expense at the manufacturers' feet is not unfair

²³⁰ *Id.*

²³¹ The FDA requires drug manufacturers to test all new products for safety and submit the results of all studies to the agency prior to approval. 21 U.S.C. § 355 (1983); H. GRABOWSKI & J. BERNON, *THE REGULATION OF PHARMACEUTICALS, BALANCING THE BENEFITS AND RISKS* 4 (1983).

²³² The Center for Disease Control ("CDC") conducted studies on Agent Orange and the swine-flu epidemic. See Malcolm Gladwell, *U.S. Firms Abandoning Birth Control Industry in Wake of Lawsuits*, WASH. POST, May 1, 1988, at J1. "At its peak use, an estimated 20% to 25% of pregnant women in the United States used Bendectin." Scolnick, *supra* note 52.

²³³ In the case of Bendectin, it might also be noted that no more epidemiological research is forthcoming, as it has been withdrawn from the market in this country. Thus plaintiffs would have nothing to gain by waiting.

²³⁴ Sustained litigation over a drug alleged to be harmful tends to affect an increase in the amount of scientific research conducted on the drug. The Bendectin litigation provided a clear example of this phenomenon. Sanders, *supra* note 192, at 346 ("The volume and sophistication of studies focusing specifically on Bendectin was, in large part, the result of the litigation.").

or unreasonable. Manufacturers should bear the expense of ascertaining the safety of products from which they profit. They are also better able to distribute the cost because they set prices. The FDA already requires manufacturers to test their own products and provide the agency with their results.²³⁵ The FDA has been criticized for allowing drug manufacturers to rush products to the market without adequate testing for safety. Allowing preliminary epidemiological evidence into the courtroom would help toxic tort litigation serve as a check on the efficacy of the FDA.²³⁶

CONCLUSION

Courts should follow the standards set by epidemiologists in determining causation, unless the causation issue stands at the "frontier" of epidemiological inquiry. If the causation issue is in the frontier, courts should admit the currently available epidemiological evidence and allow expert testimony on causation to provide for as fair a trial as possible given the lack of definitive research on the issue. This will encourage drug manufacturers to provide for adequate research in ascertaining the safety of their products, at least when that safety is called into question.

This proposed standard takes into account the factors that make legal inquiry different from scientific inquiry.²³⁷ It encourages courts to analyze the reasoning behind use of the .05 significance level in scientific research. Such a standard also allows courts to decide the conditions under which this scientific convention is inappropriate and to better tailor the fact finding inquiry into the circumstances of the case. The proposed standard concerns *only* admissibility; any problems with inferring causation from a standard less than .05 significance can be dealt with by attacks on the weight of the evidence during cross-examination.²³⁸

²³⁵ See *supra* note 230.

²³⁶ The FDA has, on occasion, failed to root out harmful drugs from the market, as was the case with D.E.S. For other examples, see Sanders, *supra* note 192, at 312-16.

²³⁷ See Cohen, *supra* note 39, at 412; Kaye, *supra* note 146, at 20; Nesson, *supra* note 47, at 528-30.

²³⁸ See McCORMICK, *supra* note 27, at 650; *United States v. Downing*, 736 F.2d 1224 (3d Cir. 1985).

Even if admitted, the epidemiological evidence may nonetheless be found insufficient to defeat a motion for summary judgment or j.n.o.v. See *Marder v. G.D. Searle*, 630

The Third Circuit's adoption of the *Downing* test in *DeLuca* was a major improvement over the use of the *Ferebee* and *Frye* standards by other Bendectin courts. However, the Third Circuit in *DeLuca* failed to give sufficient guidance to the district court on *when* experts may testify to causation based on non-significant epidemiological evidence. Since the admissibility of novel causation evidence without statistical significance continues to be an issue, the time is ripe to establish a workable standard.²³⁹

Constantine Kokkoris

EDITOR'S NOTE

On October 13, 1992 the United States Supreme Court agreed to hear an appeal of the Ninth Circuit case *Daubert v. Merrell Dow Pharmaceuticals*.²⁴⁰ The issue is whether Congress's adoption of Federal Rule of Evidence 702 supersedes the judicially created *Frye* rule.²⁴¹ The Court will consider how widely accepted a scientific theory must be for it to be admissible as evidence in federal courts.²⁴²

F. Supp. 1087, 1093-94 (D.Md. 1986) ("Rule 703, Fed. R. Evid. was intended to broaden the acceptable bases of expert opinion and to enable reliable evidence to be admitted. Admitting such evidence does not preclude the possibility of removing the case from the jury at a later stage if it is determined that in its totality the evidence was insufficient.") *Id.* at 1094 (citing *Merit Motors, Inc. v. Chrysler*, 569 F.2d 666, 673 (D.C. Cir. 1977)). In *Marder* the court "erred on the side of admissibility to allow the jury access to the maximum amount of evidence available." *Marder*, 630 F. Supp. at 1093-94.

²³⁹ For example, there is nationwide litigation pending over the drug Prozac. Prozac is a drug prescribed to psychiatric patients to combat depression. It is alleged that the drug causes suicidal behavior in its users, but none of the studies done on it have attained statistical significance. Once again, the problem is that there is a very low incidence rate of suicide, even for those taking the drug. See Paula Span, *The Man Behind the Bitter Pill Debate*, WASH. POST, Aug. 14, 1991, at C1; Natalie Angier, *Eli Lilly Facing Million-Dollar Suits on its Antidepressant Drug Prozac*, N.Y. TIMES, Aug. 16, 1990, at B13.

²⁴⁰ 951 F.2d 1128 (9th Cir. 1991), *cert. granted*, 61 U.S.L.W. 3061 (U.S. Oct. 13, 1992) (No. 92-102). For a discussion of *Daubert*, see *supra* notes 135-36 and accompanying text. For a review of Federal Rule of Evidence 702, see *supra* notes 167-68, 171 and accompanying text.

²⁴¹ See Linda Greenhouse, *High Court to Rule on Admissibility of Scientific Evidence in U.S. Courts*, N.Y. TIMES, Oct. 14, 1992, at A16; *supra* note 136.

²⁴² Greenhouse, *supra* note 241, at A16.

