

Daubert v. Merrell Dow Pharmaceuticals: Active Judicial Scrutiny of Scientific Evidence

EDWARD W. KIRSCH*

I. INTRODUCTION

The Supreme Court's opinion in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*¹ replaces the long-standing *Frye* test² governing the admissibility of scientific evidence with a new standard that places a premium on scientific validity and reliability. The new standard, as articulated by the Court, is open-ended and increases the discretion of trial judges with respect to the admission of scientific evidence. The standard fails, however, to distinguish adequately between the scientific concept of validity and the legal question of reliability, and therefore may generate confusion and inconsistent results until it receives additional judicial attention. Nonetheless, the *Daubert* standard represents a significant improvement over the *Frye* test and the reasonable medical certainty test³ previously utilized to scrutinize expert testimony in toxic tort cases. The standard focuses judicial attention correctly on an often-neglected issue in toxic tort litigation: whether the proffered testimony is based on scientifically valid reasoning connecting facts to conclusions. The inquiry required by the test will require judges to scrutinize scientific studies and evidence with which they are uncomfortable.⁴ This is preferable to allowing a case to culminate in a battle of the experts with no referee, in which juries purport to rely on science, but actually base their conclusions on dubious reasoning that are at odds with the views of the scientific community as expressed in the literature.

The deficiencies of the *Frye* and reasonable medical certainty standards are apparent in two recent cases: *Ferebee v. Chevron Chemical Co.*⁵ and *Wells v. Ortho Pharmaceutical Corp.*⁶ These cases avoided the issue of scientific validity and left juries to speculate and reach conclusions that did not comport with the state of scientific knowledge. Once it is fully understood by the lower courts, the *Daubert* standard can provide an effective means to screen evidence for conformance to the scientific method and accepted scientific practice. The *Daubert* standard should result in the admission of only scientifically sound evidence leading to decisions more consistent with science and with each other. Ironically, despite the liberal thrust of the federal rules of evidence emphasized by the Court in establishing the new standard, the test, with its focus on

* Mr. Kirsch is a law student at the George Washington University National Law Center. The author would like to acknowledge Professor Teresa Schwartz for her invaluable assistance with this article.

¹ 113 S. Ct. 2786, 2795 (1993).

² This test was articulated by the D.C. Circuit Court and had stood for 70 years. *Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923).

³ Courts using the reasonable medical certainty standard include: *Johnson v. United States*, 597 F. Supp. 374, 412 (D. Kan. 1984), *cert. denied*, 108 S. Ct. 694 (1988); *Commonwealth v. Hart*, 801 A.2d 675, 677 (Pa. 1985).

⁴ The Federal Judicial Center, the research arm of the federal courts, published a reference manual in 1994, to assist judges in managing cases involving scientific evidence by providing an overview of seven commonly encountered types of scientific evidence. FEDERAL JUDICIAL CENTER, REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 3 (1994).

⁵ 736 F.2d 1529 (D.C. Cir.), *cert. denied*, 469 U.S. 1062 (1984).

⁶ 788 F.2d 741 (11th Cir.), *cert. denied*, 479 U.S. 950 (1986).

scientific validity, will provide for the liberal admission of scientific evidence, but also will invite close judicial scrutiny of the causation evidence often offered in toxic tort litigation, resulting in an increased tendency of courts to direct a verdict or grant summary judgment.

II. THE NEW STANDARD

A. *The Distinction Between Validity and Reliability Captured in Daubert?*

Daubert supplants the long-standing *Frye* test governing the admissibility of scientific evidence by placing a premium on scientific validity and reliability. Scientists frequently draw a distinction between the related concepts of scientific validity and reliability.⁷ This distinction is crucial to a full understanding of the Supreme Court's opinion in *Daubert*, the deficiencies of the *Frye* standard, and the potential of the *Daubert* standard to provide an effective means to screen scientific evidence for conformance to the scientific method and accepted scientific practice.

The Supreme Court in *Daubert* acknowledged the importance of the distinction between scientific validity and reliability. The Court then distinguished between the two concepts similarly to the commentator, Bert Black.⁸ In Black's view, "reliability means that a successful outcome, or a correct answer, is sufficiently probable for a given situation."⁹ For example, "[a] baseball player who gets a hit forty percent of the time is an extremely reliable batter, but a lie detection device that correctly indicates falsehood ninety percent of the time may not be reliable enough for use at trial."¹⁰ Scientific validity on the other hand, according to Black, refers to "that which results from sound and cogent reasoning."¹¹ Validity addresses the issue of whether a technique is grounded firmly in scientific principles such that it is capable of producing the conclusions sought. An inquiry into scientific validity focuses on methodology and examines the reasoning that leads to a conclusion; a reliability inquiry focuses on the predictive power of a conclusion, or the inherent potential error rate of a technique, and the procedural care taken to ensure accuracy. Whether the reliability of a technique such as a polygraph test is sufficient to warrant the admission of its results into evidence is purely a legal determination based on an assessment of scientific reliability.¹² However, when the pattern of reasoning underlying a technique is called into doubt, analysis must be conducted to determine the validity of the principles governing the test or technology. The inquiry is necessary because, as Black notes, "[a]n invalid conclusion cannot be reliable, yet valid reasoning does not necessarily lead to reliable conclusions."¹³ It is crucial that courts examine the scientific validity of forms of evidence that are based on controversial reasoning. Forensic tests such as spectrographic voice analysis,¹⁴ and evi-

⁷ See, e.g., Bert Black, *A Unified Theory of Scientific Evidence*, 56 *FORDHAM L. REV.* 595, 599 (1988).

⁸ See Black, *supra* note 7, at 599.

⁹ *Id.*

¹⁰ *Id.*

¹¹ *Id.*

¹² *Id.*

¹³ *Id.*

¹⁴ Spectrographic voice analysis is a technique in which voice samples obtained from a suspect are compared to samples, usually tape recordings, of the perpetrator's voice. An attempt is made to obtain a match by comparing the patterns generated by a spectrograph. Spectrographic voice analysis rests on the

dence of causation that is often introduced in toxic tort litigation, such as animal studies, raise validity and reliability issues.¹⁵ The scientific validity issues, however, often have been overlooked under the *Frye* regime, which tends to obscure the inquiry. This article will focus on the implications of the *Daubert* decision for the admissibility of evidence in toxic tort litigation, although the Court's opinion addresses all scientific evidence.¹⁶

The distinction between scientific validity and reliability is crucial because it elucidates the line between scientific and legal questions.¹⁷ Bert Black views "the scientific question as a matter of validity, with the answer depending on accepted scientific practice and the soundness and cogency of the entire pattern of reasoning leading to the expert's conclusion."¹⁸ The legal question, however, is one of reliability, with the outcome turning on legal standards such as the preponderance of the evidence standard.¹⁹ *Daubert* seems to recognize the distinction between scientific validity and reliability. The Court occasionally uses the terms scientific validity and reliability interchangeably in its opinion, but seems in all instances to be referring to both the reasoning that supports a conclusion as well as the reliability of the conclusion.²⁰ Validity can be viewed as a subissue of reliability because without validity there can be no reliability. Conversely, the empirical reliability of a technique and its predictive power can serve as an indicator of validity even though the mechanisms in operation are not well understood.²¹

unproven assertion that every human voice is unique. Most of the federal circuits have admitted spectrographic voice analysis into evidence without considering the scientific validity of the technique. Their inquiries have focused on reliability and general acceptance. See, Note, *The Voice Print Dilemma: Should Voices Be Seen and Not Heard?*, 35 MD. L. REV. 267, 271 (1975).

¹⁵ Toxic tort litigation often hinges on the issue of causation. Because epidemiological studies are expensive and time consuming, plaintiffs often are forced to rely upon government funded animal studies in order to prove causation. These studies have inherent weaknesses. One problem in using animal studies to identify substances that are toxic to humans is that human cells often respond to chemical exposures differently than other species. This risk of an erroneous finding of toxicity due to interspecies variation is known as the problem of "external validity." Michael D. Green, *Expert Witnesses and Sufficiency of Evidence in Toxic Substances Litigation: The Legacy of Agent Orange and Bendectin Litigation*, 86 NW. U. L. REV. 643, 654 (1992). Additionally, many scientist are critical of the practice of subjecting the animals used in these studies to the maximum tolerated dose (MTD) of the chemical of interest. The MTD is the maximum dose a species can ingest without the onset of death from poisoning. Researchers typically expose the animals to the MTD, observe the effect on the animals, and then extrapolate from these high dose exposures to estimate the risk to humans at the much lower doses characteristic of routine human activity. A linear dose-response curve most often is utilized to infer the impact on humans, but may be inaccurate. *Id.* at 655-56. In fact, "[so] much evidence has accumulated that chemicals frequently have wholly different effects in animals and humans that officials throughout government and industry often do not act on the studies' findings" of toxicity. Joel Brinkley, *Many Say Lab-Animal Tests Fail to Measure Human Risk*, N.Y. TIMES, Mar. 23, 1993, at A1. Finally, the animals commonly are interbred in order to produce a proclivity toward tumor development, especially liver cancer. Despite these weaknesses, animal studies often provide insight into pathological mechanisms that assist scientists "in framing hypotheses and in developing study designs for epidemiological studies." FEDERAL JUDICIAL CENTER, *supra* note 4, at 130.

¹⁶ A toxic tort in general is "harm . . . caused by exposure to a substance that increases the risk of contracting a serious disease, but generally involve[s] a period of latency . . . prior to the onset of the disease." Black, *supra* note 7, at 602 n.29. Exposure to toxic substances can be widespread in the modern consumer age. For example, it is estimated that 17,500,000 pregnant women were exposed to the anti-nausea drug Bendectin between 1957 and 1982. *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 43 F.3d 1311, 1313 (9th Cir. 1995).

¹⁷ Black, *supra* note 7, at 600.

¹⁸ *Id.*

¹⁹ *Id.*

²⁰ See *infra* notes 27-32 and accompanying text.

²¹ Black, *supra* note 7, at 613.

Although the Court's interchangeable use of the terms often obscures an important distinction, it is not an incorrect use of the concepts as understood by scientists.

B. *The Court's Statutory Interpretation in Daubert and the Emergence of the Scientific Validity Standard*

In *Daubert* two children and their parents sued the manufacturer of Bendectin, an extensively marketed anti-nausea drug, claiming that ingestion of the drug during pregnancy caused the severe birth defects suffered by the children.²² The plaintiffs attempted to establish causation by introducing testimony based on animal studies, chemical structure analysis, and re-analysis of previously published epidemiological studies.²³ The district court ruled this evidence inadmissible given the overwhelming epidemiological evidence refuting the contention that Bendectin is a teratogen.²⁴ The Ninth Circuit affirmed the lower courts decision to not admit the evidence and granted the defendant's motion for summary judgment.²⁵ The Ninth Circuit affirmed the exclusion of the re-analysis evidence on the grounds that such evidence could be accepted under the *Frye* standard "only when it is subjected to verification and scrutiny by others in the field" in the normal peer review process.²⁶

The Supreme Court rejected the lower court's analysis, holding that the general acceptance test is no longer "a necessary precondition to the admissibility of scientific evidence under the Federal Rules of Evidence."²⁷ The Court determined that the Federal Rules have superseded the *Frye* test, and that they require judges to play an active gatekeeping role with respect to admission of scientific evidence under a new standard.²⁸

The Supreme Court focused on Rule 702 of the Federal Rules of Evidence governing expert testimony, which 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."²⁹ The Court determined that Rule 702 requires the judge to apply a two-part test to determine the admissibility of scientific evidence. This test requires a threshold assessment to "ensure that any and all scientific testimony or evidence admitted is not only relevant, but reliable."³⁰ The Court substantiated a scientific validity requirement in Rule 702 by noting that "[t]he adjective 'scientific' implies a grounding in the methods and procedures of science" and "the word 'knowledge' connotes more than subjective belief or unsupported

²² 113 S. Ct. at 2791.

²³ *Id.* at 2791-92.

²⁴ The trial court noted that over 30 epidemiological studies involving over 130,000 patients were introduced at trial and found no relationship between Bendectin and birth defects, indicating that "the strongest inference to be drawn for the plaintiffs . . . is that Bendectin could *possibly* have caused plaintiffs' injuries, therefore summary judgment is proper against them." 727 F. Supp. 570, 576 (S.D. Cal. 1989).

²⁵ 951 F.2d 1128, 1131 (9th Cir. 1991).

²⁶ *Id.*

²⁷ 113 S. Ct. at 2799.

²⁸ *Id.* at 2793-95. Applying the *Daubert* test on remand, the Ninth Circuit excluded as "personal opinion, not science" the testimony of a plaintiff's expert (Dr. Palmer) who had examined the plaintiff's medical records. 43 F.3d at 1319.

²⁹ FED. R. EVID. 702.

³⁰ 113 S. Ct. at 2795.

speculation.”³¹ The Court then asserted that “in order to qualify as ‘scientific knowledge,’ an inference or assertion must be derived by the scientific method.”³²

The Court somewhat confusingly continued to describe this threshold requirement as one of “evidentiary reliability.”³³ The Court clarified its terminology by quoting Black’s distinction between “reliability” and “validity,” and concluding “that in a case involving scientific evidence, *evidentiary reliability* will be based on *scientific validity*.”³⁴ The Court apparently recognized that because there can be no reliability without validity, validity can be viewed as a sub-issue of reliability. Throughout the opinion the Court used these terms interchangeably, obscuring the distinction, but at all times seeming to require a showing of the preconditions of validity and reliability, and the independent requirement of relevance.³⁵

The relevance requirement is not controversial and is derived from Rule 702, which requires evidence to “assist the trier of fact . . . to determine a fact in issue.”³⁶ The relevancy component of the standard “requires a scientifically valid connection to the pertinent inquiry” or a “fit” with an issue in the case.³⁷ The requirement is added because “scientific validity for one purpose is not necessarily scientific validity for other, unrelated purposes.”³⁸

The Supreme Court intended to provide judges with an effective means to screen junk science through the application of the new standard. It evidenced this concern by taking issue with the contention of amici that rejection of the *Frye* test would result in admission of “irrational pseudoscientific assertions.”³⁹ By placing the focus of analysis on the scientific validity, reliability, and relevance of proffered scientific evidence, the new standard should prove more effective in screening junk science than the general acceptance standard without incurring its flaws.⁴⁰ The *Daubert* test forces courts to evaluate the validity of the reasoning process that an expert uses to connect data to conclusions resulting in decisions that should be consistent from court to court and comport with the status of scientific knowledge.⁴¹

C. The *Daubert* Factors: An Early Interpretation

After setting forth the general terms of the new standard, the *Daubert* Court attempted to elucidate the standard by providing a list of factors that would bear on the scientific validity and reliability inquiry.⁴² While the Court wisely “[did] not presume to

³¹ *Id.*

³² *Id.*

³³ *Id.* n.9.

³⁴ *Id.* (emphasis added).

³⁵ The Court, for example, summarizes the *Daubert* test as requiring “a preliminary assessment of whether the reasoning or methodology underlying the testimony is scientifically valid and . . . can be applied to the facts in issue.” *Id.* at 2796. In this passage, no mention is made of scientific reliability. It is obvious, however, from the court provided list of factor that bear on the inquiry, that a theory or technique must be determined to be scientifically valid and reliable. In particular, the error rate and the existence and maintenance of standards are factors that pertain to reliability rather than validity. *Id.* at 2797.

³⁶ *Id.* at 2795. Rule 402 establishes a presumption that all relevant evidence is admissible. FED. R. EVID. 402.

³⁷ 113 S. Ct. at 2796.

³⁸ *Id.*

³⁹ *Id.* at 2798.

⁴⁰ The deficiencies of the *Frye* general acceptance test are explored *infra* part III of this article.

⁴¹ See *infra* text accompanying notes 122-86.

⁴² 113 S. Ct. at 2796.

set out a definitive checklist or test," it did provide five factors that can be employed to analyze evidence in many contexts:⁴³ whether the theory or technique "can (and has been) tested"; whether it has been "subjected to peer review and publication"; whether the potential rate of error indicates it is reliable; whether standards have been observed which control the technique's operation; and, finally, whether the theory or technique is "generally accepted" by the scientific community.⁴⁴ Thus, the *Frye* test may serve as an indication of the scientific validity, but not as a dispositive test in its own right.

Three of these factors, peer review, general acceptance and testability, are integral components of modern scientific inquiry. Therefore, they can serve as indicators of scientific validity and reliability for most types of proffered scientific evidence. The Court seems to be aware however, that it is impossible, or at least unwise, to try to establish an exhaustive list of validity and reliability criteria as they necessarily must vary with the type of evidence presented. For example, the relevant indicators of validity and reliability are different between spectrographic voice evidence, where the potential rate of error under field conditions is a dominant consideration, and epidemiological evidence, where members of the field frequently utilize five sufficiency criteria and the relative risk ratio to assess this type of scientific evidence.⁴⁵

One of the few cases that has applied the *Daubert* standard in a toxic tort setting illustrates the flexibility of the standard and the need to use evaluation factors appropriate to the type of evidence offered. In *In re Joint Eastern and Southern District Asbestos Litigation*,⁴⁶ the district court found it necessary to use the relative risk ratio and five sufficiency criteria commonly used by epidemiologists to assess the sufficiency of the epidemiological evidence offered by the plaintiff. Plaintiff's theory for recovery in the case was that a group of sheet metal workers engaged in building construction, including the plaintiff Maiorana, were exposed to an asbestos-containing fireproof spray, Cafco D, which proximately caused his colon cancer.⁴⁷ Defendants moved for a judgment notwithstanding the verdict⁴⁸ after a jury verdict for the plaintiff.⁴⁹ The court granted the motion after examining the validity of the plaintiff's proffered evidence against the five sufficiency criteria, and concluded that the evidence was insufficient to create an issue of fact for the jury to decide.⁵⁰

The court conducted its inquiry into the sufficiency of the plaintiff's evidence by first noting that epidemiologists begin their analysis by determining if a relationship

⁴³ *Id.*

⁴⁴ *Id.* at 2797.

⁴⁵ See *United States v. Williams*, 583 F.2d 1194 (2d Cir. 1978), *cert. denied*, 439 U.S. 1117 (1979). The most extensive experiments to investigate the reliability of spectrographic voice identification were conducted by Dr. Oscar Tosi and completed in 1970. These controlled experiments demonstrated that under laboratory conditions the rate of false identifications for the technique was no greater than 9.8% and could be reduced with improved training of technicians. R.H. Bolt et al., *Speaker Identification by Speech Spectrograms: Some Further Observations*, 54 J. ACOUSTICAL SOC'Y AM. 531, 532 (1974). In 1979, a multidisciplinary team comprising scientists in the fields of acoustics, physics, speech, and audiology, and sponsored by the National Academy of Sciences, reviewed the scientific basis of spectrography and the results of Dr. Tosi's experiment and published a highly critical assessment. The committee concluded that the "5% to 10% false-identification rates" extolled by Tosi were "artificial minima which are likely to increase when conditions depart from the laboratory situation" of the experiment. R.H. Bolt et al., *Speaker Identification by Speech Spectrograms: Some Further Observations*, 54 J. ACOUSTICAL SOC'Y AM. 531, 533 (1974). See *infra* note 58 for a discussion of the sufficiency criteria used to assess the results of an epidemiological study.

⁴⁶ 827 F. Supp. 1014, 1037 (S.D.N.Y. 1993).

⁴⁷ *Id.* at 1024-25.

⁴⁸ FED. R. CIV. P. 50(b).

⁴⁹ 827 F. Supp. at 1023.

⁵⁰ *Id.* at 1037-50.

between exposure to a substance and a malady is “biologically possible.”⁵¹ Epidemiologists then seek to determine whether a statistically significant causal relationship exists between the exposure of a cohort group to a substance and occurrence of the malady.⁵² The strength of an association is measured by epidemiologists using the relative risk ratio, also known in studies of asbestos as the standard mortality ratio (SMR). “A relative risk of 2.0 means that, on the average, there is a fifty percent likelihood that a particular case of the disease was caused by the event under investigation and a fifty percent likelihood that the disease was caused by chance alone.”⁵³ “A relative risk greater than 2.0 means that the disease more likely than not was caused by the event [exposure to the suspect chemical].”⁵⁴

Because scientists have limited understanding of the mechanism by which chemicals cause cancer, standard scientific practice requires the epidemiologist to conduct additional analysis using five sufficiency criteria.⁵⁵ The district court recognized this common practice of the scientific community by noting that in addition to the relative risk ratio, five sufficiency criteria “must be analyzed to determine whether or not evidence has achieved the status of scientific knowledge.”⁵⁶ The five sufficiency criteria recognized by the district court, and scientists, include consistency of association, dose-response relationship, results of experimental studies such as animal studies, plausibility of a causal link, and coherence.⁵⁷ The court adeptly applied the five sufficiency criteria to the evidence offered in the case and found that plaintiff’s evidence could not be relied on because it failed to satisfy any of these criteria.⁵⁸

Joint Eastern and Southern District Asbestos illustrates that beyond the court’s focus on scientific validity in addressing the admissibility of scientific evidence, the *Daubert* standard is a flexible one that must be tailored and requires elaboration to fit the type of evidence under consideration. *Joint Eastern* demonstrates that this challenge is within the competence of a court. In determining whether the plaintiff’s evi-

⁵¹ *Id.* at 1027.

⁵² *Id.* “There are two main types of observational studies: cohort studies and case-control studies.” FEDERAL JUDICIAL CENTER, *supra* note 4, at 134. In the present case, the court examined the result of cohort studies. “Cohort studies, which use exposure as an independent variable, compare two groups: one group that is exposed to the agent, and a control group that consists of persons with similar characteristics who have not been exposed.” *Id.*

⁵³ 827 F. Supp. at 1027.

⁵⁴ *Id.*

⁵⁵ *Id.* at 1028.

⁵⁶ *Id.* at 1037.

⁵⁷ *Id.* at 1037-38.

⁵⁸ *Id.* at 1038.

The five factors include: first, the consistency of the association between [exposure] and [colon cancer], which raises the question, Is the SMR of a single epidemiological study addressing the relationship between [exposure] and [disease] consistent with the SMRs derived in other epidemiological studies?; second, the dose-response relationship between [exposure] and [disease]: What is the epidemiological response in a cohort to estimated doses of [asbestos]?; third, the results of experimental studies: Have experimental studies been conducted animals, for example, and if so, were they positive?; fourth, the plausibility of there being a biological link between [asbestos] and [colon cancer]: Given the biological and chemical mechanisms involved, what is the degree of probability that exposure to [asbestos] can give rise to the subsequent development of [colon cancer]?; and fifth, the coherence between [exposure] and [malady]: How many “confounding or alternative factors or conditions can contribute to the development of [colon cancer] in A? And how difficult is it to exclude these confounding conditions thereby isolating A’s exposure to [asbestos] as the statistically significant and ‘more likely than not’ cause of [colon cancer] in A?”

Id. at 1037. These factors can be found in A Bradford Hill, *The Environment and Disease: Association or Causation*, 58 PROC. ROYAL SOC’Y MED. 295, 295-300 (1965).

dence met the threshold scientific validity requirement, the district court did not limit itself to the five criteria outlined in the *Daubert* case. Instead, the court found it necessary to examine the “standardized mortality ratio (SMR) which defines the strength of a possible association, a well as five ‘sufficiency criteria’” commonly utilized by epidemiologists to draw conclusions from epidemiological evidence.⁵⁹ The district court in this case proved effective in identifying the appropriate indicators of validity and reliability, and in applying them to the proffered evidence in the case. The result was that the plaintiff’s weak causation evidence was found inadequate to carry the burden of proof. This may be a common outcome under the *Daubert* standard. Although the scientific validity standard will assist plaintiffs relying on weak causation evidence by liberally allowing the admission of scientific evidence, the standard is likely to prompt an active judicial review of the propriety of submitting the causation issue to the jury as well as more frequent use of summary judgment and other devices to dismiss toxic tort cases.

D. The New Standard has Wide Application

While the focus of this article is on the impact of the *Daubert* decision on toxic tort litigation, the admissibility of scientific evidence is a critical issue in criminal cases and other controversies. The new standard will be used to examine DNA profiling evidence, blood spatter evidence, spectrographic voice analysis, and other types of forensic evidence. The test will be applied far more often in the forensic science context due to the large volume of criminal cases relative to toxic tort cases. In addition, it is possible the Court’s interpretation of Rule 702 in *Daubert* will be applied to other forms of technical evidence such as crash-worthiness evidence in product liability cases. Justice Blackmun notes in the opinion that “Rule 702 also applies to ‘technical, or other specialized knowledge,’” but chose to limit his discussion to the scientific type of evidence at issue in the case at bar.⁶⁰

III. THE *Frye* AND REASONABLE MEDICAL CERTAINTY TEST LEGACY

A. The Deficiencies of the Frye Standard

1. Identifying the Relevant Field and the Level of General Acceptance

In order to appreciate the impact of the new standard set forth in *Daubert*, it is necessary to examine the deficiencies of its predecessors. The *Frye* test was first established in a case⁶¹ that the Supreme Court aptly described as “a short citation-free 1923 decision” that “concern[ed] the admissibility of evidence derived from . . . a crude precursor to the polygraph machine.”⁶² The test held wide judicial acceptance for sixty years and was set forth as follows:

⁵⁹ 827 F. Supp. at 1038.

⁶⁰ 113 S. Ct. at 2795 n.8.

⁶¹ 293 F. at 1014.

⁶² 113 S. Ct. at 2793.

Just when a scientific principle or discovery crosses a line between experimental and demonstrable stages is difficult to define. Somewhere in this twilight zone the evidential forces of the principle must be recognized, and while courts will go a long way in admitting expert testimony deduced from a well-recognized principle or discovery, *the thing from which the deduction must be made* must be sufficiently established to have *gained general acceptance in a particular field* in which it belongs.⁶³

Courts and commentators had been increasingly critical of the *Frye* test in recent decades, alleging that the test was vague and had a chilling effect on the admission of valuable evidence from new techniques of scientific investigation.⁶⁴ The *Frye* test was vague in several respects. First, the test did not define criteria to identify the pertinent scientific community or the level of agreement needed to establish “*general acceptance in a particular field*.”⁶⁵ As a result, courts considering the same evidence have “varied widely as to who[m] and how large the pertinent scientific community must be.”⁶⁶

In the context of spectrographic evidence, this deficiency has introduced an element of arbitrariness where the outcome of judicial decisions has become sensitive to the selection of the relevant scientific community. The Supreme Judicial Court of Massachusetts in *Commonwealth v. Lykus*, for example, held that spectrography generally was accepted by the scientific community narrowly defined to consist of “those who would be expected to be familiar with its use.”⁶⁷ The *Lykus* court’s definition of the relevant scientific community as a practical matter limited the scientific community to forensic scientists with a vested interest in the acceptance of the technique. Other courts have manipulated the ambiguity of the *Frye* test similarly in order to justify the admission or exclusion of scientific evidence. In one case, the court admitted that the use of a Nalline technique to determine narcotics use generally was not accepted by the medical profession, but found it sufficient that the test was accepted by forensic scientists.⁶⁸

There exists a substantial risk of bootstrapping or self-validation when the relevant scientific community is limited to forensic scientists specializing in the use of a technique as occurred in the above mentioned cases. In the spectrography cases, courts inappropriately relied on the founders and initial proponents of the technology whose professional reputations often depended on the acceptance of the technique. The Supreme Court of New Jersey recognized the problem and voiced its concern noting that spectrography “appears to be a sole source industry, with the principal place of research located at the Michigan State University Speech and Hearing Research Lab.”⁶⁹ The New Jersey court’s concern was that the experts who most often testified that the use of

⁶³ 293 F. at 1014 (emphasis added).

⁶⁴ *United States v. Downing*, 753 F.2d 1224, 1236 (3d. Cir. 1985). The Supreme Court relied on the precedent of *United States v. Downing*, in rejecting the general acceptance test. The Third Circuit Court of Appeals in *Downing* derived a test from the helpfulness standard of Federal Rule of Evidence 702 in which the reliability and “fit” of the proffered evidence was to be balanced against the tendency to mislead the jury. *Id.* at 1226. In language later echoed in *Daubert*, the *Downing* court focused its reliability inquiry on the “principles upon which the expert testimony rests” and held that general acceptance could serve as an indicator of reliability but should not serve as an independent test. *Id.* at 1238.

⁶⁵ *Id.* at 1236.

⁶⁶ John F. Decker & Joel Handler, *Voiceprint Identification Evidence—Out of the Frye Pan and into Admissibility*, 26 AM. U. L. REV. 314, 362 (1977).

⁶⁷ 327 N.E.2d 671, 677 (Mass. 1975).

⁶⁸ *People v. Williams*, 331 P.2d 251 (Cal. App. Dep’t Super. Ct. 1958).

⁶⁹ *Windmere v. International Ins. Co.*, 522 A.2d 405, 408 (N.J. 1987).

spectrographs to identify human voices generally was accepted were those who most closely identified with the development of the technology and had built their careers on spectrographic research.⁷⁰

In contrast to the decisions above, a federal court of appeals took a broad view of the pertinent scientific community in finding spectrography “was not sufficiently accepted by the scientific community as a whole” and was inadmissible.⁷¹ Other federal courts of appeals have examined the same empirical studies and heard the testimony of many of the same expert witnesses, yet reached the conclusion that the test generally was accepted by narrowing their definition of the relevant scientific community, as did the *Lykus* court.⁷²

2. Identifying a “Novel” Technique

The *Frye* court applied the general acceptance test to what it considered to be a novel technique, prompting later courts to limit the applicability of the *Frye* standard to new scientific techniques.⁷³ This practice has resulted in a second area of ambiguity because the concept of a novel technique has proven to be vague and malleable. Various courts purporting to adhere to the *Frye* standard have avoided the rigors of the *Frye* test by construing a given technique to be in the mainstream of science.⁷⁴ Conversely, other courts have deemed the same evidence “novel” and have applied the exacting *Frye* standard.⁷⁵

Ferebee illustrates the selective application of the defunct *Frye* standard in a toxic torts case. In that case, the United States Court of Appeals for the District of Columbia refused to apply the *Frye* test even though it nominally adhered to the standard. The circuit court concluded, with little supporting analysis, that the *Frye* test applied only to “the introduction of evidence based on novel scientific techniques or methodologies” rather than the controversial conclusions of an expert based on a “well-founded” methodology as in the present case.⁷⁶ The court explicitly avoided reviewing the scientific validity of the reasoning of the medical expert testimony presented in the case, and chose instead to rely on the unfounded personal opinions of the plaintiff’s two medical experts that an agricultural worker’s pulmonary fibrosis was caused by long-term skin exposure to low levels of the herbicide paraquat.⁷⁷ The “well-founded” methodology employed in *Ferebee* was traditional differential diagnosis based on personal observation of the patient.⁷⁸ The court ignored the fact that this methodology has no predictive power in this context because the effects of low level exposure to paraquat are statistically unknown. Additionally, the court seemed to ignore the well-founded evidence of

⁷⁰ *Id.*

⁷¹ *United States v. Addison*, 498 F.2d 741, 742 (D.C. Cir. 1974).

⁷² *See United States v. Smith*, 869 F.2d 348, 354 (7th Cir. 1989); *United States v. Baller*, 519 F.2d 463 (4th Cir.), *cert. denied*, 96 S. Ct. 456 (1975); *United States v. Franks*, 511 F.2d 25 (6th Cir.), *cert. denied*, 95 S. Ct. 2654 (1975); *Williams*, 583 F.2d at 1194.

⁷³ 293 F. at 1014.

⁷⁴ *See, e.g., Ferebee*, 736 F.2d at 1529; *United States v. Jenkins*, 525 F.2d 819 (6th Cir. 1975) (general acceptance test not applied to voiceprint evidence); *People v. Clark*, 857 P.2d 1099, 1142 (Cal. 1993) (blood splatter evidence is not new, *Kelley/Frye* rule is therefore inapplicable).

⁷⁵ *See United States v. Brown*, 557 F.2d 541 (6th Cir. 1977) (*Frye* test revived to exclude ion microprobiotic analysis); *Addison*, 498 F.2d at 743 (general acceptance test used to assess “new methods” including voiceprint evidence).

⁷⁶ 736 F.2d at 1535.

⁷⁷ *Id.* at 1532.

⁷⁸ *Id.* at 1533.

the defendant that paraquat is known only to be “acutely toxic—that is, any injuries resulting from exposure to paraquat occur within a very short time of exposure, such as days or weeks, and that when exposure ceases, so too does the injury.”⁷⁹ The plaintiff’s medical history did not fit this established pattern.⁸⁰ He exhibited his first symptoms of lung disease ten months after his last low-level exposure to the herbicide and his condition continued to deteriorate long after exposure had ceased.⁸¹ The case illustrates a point made by the Supreme Court that “scientific validity [of a methodology] for one purpose is not necessarily scientific validity for other, unrelated purposes.”⁸²

In contrast to *Ferebee*, the District Columbia Court of Appeals concluded in *United States v. Addison* that spectrographic voice analysis was a “new method of scientific measurement” that merited the heightened scrutiny of the *Frye* standard.⁸³ The court extended its inquiry beyond the demands of general acceptance and touched on scientific reliability in its criticism of the Tosi Study relied on by the prosecution.⁸⁴ The Tosi Study attempted to quantify the error rate associated with identification of tape recorded voices through the use of spectrographic analysis.⁸⁵ The Tosi Study concluded that the maximum expected error rate for false voice identifications under laboratory conditions was about 9.8%.⁸⁶ The *Addison* court, in scrutinizing these findings, correctly noted that the study “did not necessarily indicate that spectrographic analysis would enjoy a comparable success rate when applied to the general populace.”⁸⁷ In reaching this conclusion the court conducted only a perfunctory analysis of the variance between the laboratory conditions utilized in the controlled study and the more challenging environment of forensic field conditions. The court avoided issues of scientific validity because such an inquiry would have required an examination of the fundamental assumptions of the technique that inter-speaker variability in speech greatly exceeds intra-speaker variability, and that no two voices are exactly alike. The court concluded that spectrographic voice evidence should not be admitted.⁸⁸ Taken together these two cases illustrate the uneven application of the *Frye* standard by a court and the malleability of the standard that results from the novel technique requirement. As exemplified by these cases, the issue of admissibility seems to hinge on whether the court deems a technique to be novel. The cases also demonstrate that under either the *Frye* or reasonable medical certainty standards, the crucial issue of scientific validity rarely will be addressed squarely.

In addition to inconsistent application and outcomes, there are other dangers inherent in limiting a review of the merits of scientific evidence to novel techniques.⁸⁹ One danger is the lack of a continuing obligation to assess the scientific merit of a technique as the body of scientific knowledge evolves over time.⁹⁰ Under *Frye*, once a novel technique or theory has met the test of general acceptance, there exists a strong probability that courts will rely on precedence and continue to admit the evidence “in spite

⁷⁹ *Id.* at 1535.

⁸⁰ Black, *supra* note 7, at 671.

⁸¹ *Id.*

⁸² *Daubert*, 113 S. Ct. at 2796.

⁸³ 498 F.2d at 743.

⁸⁴ *Id.* at 744.

⁸⁵ *Id.*

⁸⁶ *See Smith*, 869 F.2d at 353.

⁸⁷ *See* 498 F.2d at 744.

⁸⁸ *Id.* at 745.

⁸⁹ James E. Starrs, *Frye v. United States Restructured and Revitalized: A Proposal to Amend Federal Evidence Rule 702*, JURIMETRICS J., Spring 1986, at 249, 253.

⁹⁰ *Id.*

of . . . unambiguous and uniform rejection by the scientific community” at a later date.⁹¹ Science steadily evolves, so that a conclusion or technique that is generally accepted at one moment in time may later be found invalid. An example is the forensic technique of “testing for gunpowder residues through the use of dipheylamine reagent,” popularly known as the “paraffin test.”⁹² The paraffin test was employed widely for nearly thirty years “to detect gunshot residue on the hand of a person who has recently fired a weapon.”⁹³ Despite a series of studies and papers that raised doubts as to the validity of the technique over decades, the test continued to elicit wide acceptance until a study completed in 1967 conclusively “found the test to be unreliable.”⁹⁴ Because it is not limited to novel science, the scientific validity test of the *Daubert* opinion should ensure that any technique relied on by the trier of fact has an adequate foundation in light of the current state of scientific knowledge. The test should ensure that the issue of scientific validity is revisited frequently so that the legal view of a technique will evolve in concert with the body of scientific evidence.

3. *What Must Be Accepted?*

Another problem with the *Frye* test is its failure to distinguish between scientific reliability and validity. The test provides that “*the thing from which the deduction must be made*” must be “generally accepted.”⁹⁵ It is unclear whether the word “thing” refers to the conclusions resulting from the reasoning or the underlying scientific reasoning itself.⁹⁶ The reasoning connecting facts to conclusions, or scientific validity, was not disputed in the *Frye* case itself.⁹⁷ The theory that lying might cause strong emotions that prompt involuntary physiological reactions was accepted at the time of *Frye*.⁹⁸ The language of the *Frye* test neglects the distinction between scientific reliability and validity with the result that most courts applying the *Frye* standard have not addressed scientific validity. This deficiency frequently has resulted in passive admission of dubious scientific evidence with the result being a battle of experts with the jury left to speculate as to the winner. The *Frye* test as applied has resulted in inconsistent outcomes on identical evidence in toxic tort cases where the evidence presented is complex and its validity often controversial.

United States v. Williams is an example of a case in which the validity of the reasoning connecting facts to conclusions was ignored occurred.⁹⁹ The court weighed the probative value of the evidence against the tendency to mislead or prejudice the jury in deciding to admit spectrographic voice print evidence. The court did not examine the empirical evidence nor the reasoning underlying the technique; it focused almost exclusively on the reliability of the technique, and did not question the technique’s basic assumption that inter-speaker variability in speech greatly exceeds intra-speaker variability. Other assumptions that pertain to the validity of the reasoning also were not questioned, including the assumption that no two voices are exactly alike. The subjec-

⁹¹ *Id.*

⁹² *Id.*

⁹³ Paul C. Giannelli, *The Admissibility of Novel Scientific Evidence: Frye v. United States a Half Century Later*, 80 COLUM. L. REV. 1197 (1980).

⁹⁴ *Id.* at 1225.

⁹⁵ 293 F. at 1014.

⁹⁶ Black, *supra* note 7, at 602.

⁹⁷ *Id.* at 630.

⁹⁸ *Id.*

⁹⁹ 583 F.2d at 1194 (cited in Black, *supra* note 7, at 608).

tive nature of the points of similarity used to produce a match, and their lack of a firm analytic basis, also was not analyzed by any of the federal circuits that admitted the evidence under the *Frye* standard.

4. *The Test is Conservative*

Critics of the *Frye* test maintain that it is too conservative and unduly restricts the admission of new scientific evidence. The Court of Appeals for the Third Circuit rejected the *Frye* test in *Downing*, decrying its conservative approach to the admission of scientific evidence as being “at odds with the spirit, if not the precise language, of the Federal Rules of Evidence.”¹⁰⁰ According to the *Downing* court, the test resulted in the exclusion of “much probative and reliable information from the jury’s consideration, thereby unnecessarily limiting the truth-seeking function of litigation.”¹⁰¹ While this danger seems obvious, there is an equal danger that junk science will be admitted because the *Frye* test does not focus on scientific validity. This is what occurred in *Ferebee* and in numerous criminal cases that relied on the dubious “paraffin test.” The conservatism of the *Frye* test arguably is justified, however, as pointed out by Giannelli: “the critical issue is whether other approaches can better achieve the *Frye* objective of ‘preventing the introduction into evidence of specious and unfounded scientific [evidence]’” while avoiding a chilling effect on the admission of new forms of scientific evidence.¹⁰² The *Daubert* standard, properly applied, should enable courts to achieve these twin objectives by focusing on scientific validity.

B. *Inconsistent Application of the Frye Standard*

Commentators have criticized severely the *Daubert* decision to discard the general acceptance test, pointing out that the *Frye* test was “long-established” and had a “substantial body of law interpreting it” making it “relatively easy for trial judges to apply.”¹⁰³ Although the commentator is correct that the *Frye* rule was widely adopted, with ten of the thirteen federal circuits and many states using it, these federal and state courts varied widely in their application of the test. Several jurisdictions, without explicitly overruling *Frye*, had construed the test in such a way that they were conducting an inquiry into reliability, not general acceptance. The Sixth Circuit in *United States v. Franks*¹⁰⁴ utilized this approach; it claimed to apply *Frye* in the examination of spectrographic voice evidence, but essentially engaged in a reliability inquiry.¹⁰⁵ There was not a single *Frye* test; rather, due to its deficiencies, the courts effectively crafted a variety of standards that they attached to *Frye*. Several of these standards, such as the reliability standard of *Franks*, resemble the general acceptance test in form only. Despite the large body of law that attempts to interpret *Frye*, these interpretations have varied because of the test’s inherent vagueness.

¹⁰⁰ 753 F.2d at 1237.

¹⁰¹ *Id.* at 1236.

¹⁰² Giannelli, *supra* note 93, at 1224.

¹⁰³ Timothy B. Dyke & Gregory A. Castanias, *Daubert Doesn’t End Debate on Experts*, NAT’L L.J., Aug. 2, 1993, at 4.

¹⁰⁴ 511 F.2d 25 (6th Cir.), *cert. denied*, 422 U.S. 1042 (1975).

¹⁰⁵ *Id.*

C. *Inconsistent Results in Toxic Torts Cases That Defy Science*

The *Frye* test enables judges to avoid scrutiny of scientific studies and evidence with which they are uncomfortable. As a result many cases that involve complex scientific issues of causation or new techniques have degenerated into a battle of experts in which the judge fails to act as a referee. In these cases juries are left to speculate and often reach conclusions purporting to rely on science, but in reality are based on dubious reasoning and are at odds with the view of the scientific community as expressed in literature. The deficiencies of the *Frye* and reasonable medical certainty standards are apparent in *Ferebee*. The case illustrates a court's reliance on the unfounded personal opinion of a medical expert permitted by the reasonable medical certainty standard.¹⁰⁶ *Ferebee* also illustrates courts willingness under *Frye* to permit juries to resolve the battle of the experts. Judge Mikva, in affirming the district court's judgment upholding a wrongful death jury verdict, concluded:

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.¹⁰⁷

The court permitted the jury to resolve this classic "battle of the experts," noting that two expert witnesses had refuted Chevron's assertion that no evidence existed to show that paraquat could cause a chronic injury such as pulmonary fibrosis, and "the jury was entitled to believe those experts."¹⁰⁸ The court apparently was unconcerned that the jury was left to decide the outcome under these circumstances, stating, "[T]he case was a classic battle of the experts, a battle in which the jury must decide the victor."¹⁰⁹ The court also was not disturbed "that another jury might reach the opposite conclusion or that science would require more evidence before conclusively considering the causation question resolved."¹¹⁰

One of the most egregious cases decided under the prior regime of admissibility standards was *Wells v. Ortho Pharmaceutical Corp.*, a case in which a court completely ignored the overwhelming consensus of the scientific community. The key issue in *Wells* was one of causation in a toxic tort case. Specifically, the plaintiff alleged that an infant's limb deformities were caused by her mother's continued use of a spermicidal jelly for about four weeks after conception.¹¹¹ The trial court examined the epidemiological studies introduced by both parties and "found the studies to be inconclusive on the ultimate issue of whether the Product caused [the plaintiff's] birth defects."¹¹² In light of this conclusion, it would be expected that the district court would conclude that plaintiff had not carried the burden of proof on the causation issue to a preponderance of the evidence. Instead, the district court rested its decision on "the demeanor, tone, motives,

¹⁰⁶ See *supra* note 3

¹⁰⁷ 736 F.2d at 1535.

¹⁰⁸ *Id.* at 1535.

¹⁰⁹ *Id.*

¹¹⁰ *Id.*

¹¹¹ 788 F.2d at 742.

¹¹² *Id.* at 745.

biases, and interests that might have influenced each expert's opinion" and decided to award \$5,300,000 dollars in damages.¹¹³ The appellate court upheld this decision asserting that "it does not matter that the medical community might require more research and evidence before conclusively resolving the question."¹¹⁴ The medical community, however, thought its well established opinion mattered very much and, in an unusual move, chastised the *Wells* court for awarding damages to the plaintiff.¹¹⁵ Several prominent physicians wrote articles in prestigious medical journals objecting to the *Wells* decision.¹¹⁶ In one of these articles, two physicians noted that "the decision took the medical community by surprise, because the overwhelming body of evidence indicat[ed] that spermicides are not teratogenic."¹¹⁷

These cases represent extreme applications of the former admissibility standards. They demonstrate that under the *Frye*¹¹⁸ and reasonable medical certainty standards, evidence of questionable value was admitted and resulted in legal decisions that have shocked the scientific community by contradicting the body of scientific evidence. The inconsistent results reached in toxic tort litigation under the existing *Frye* standard undermine the public's opinion of the judicial system. Inconsistent outcomes exact a price on the economy as well as the legitimacy of the legal system. They have had a deterrent effect on research and development of new drugs for certain purposes, and have resulted in the removal of valuable products from the market. Merrell Dow, for example, removed Bendectin from the market in 1983 despite evidence that it probably is not harmful and its continued approval by the Food and Drug Administration.¹¹⁹ The Supreme Court in *Daubert* rejected this result and in its opinion asked other courts to engage in a "gatekeeping" role to prevent such outcomes.¹²⁰ The Court emphasized the point by asserting that "the trial judge must ensure that any and all scientific testimony or evidence admitted is not only relevant, but reliable."¹²¹

IV. DAUBERT IS THE CULMINATION OF A MOVEMENT TOWARD ACTIVE JUDICIAL REVIEW OF THE ADMISSIBILITY AND SUFFICIENCY OF SCIENTIFIC EVIDENCE

¹¹³ Black, *supra* note 7, at 673.

¹¹⁴ 788 F.2d at 745.

¹¹⁵ Black, *supra* note 7, at 673.

¹¹⁶ *Id.*

¹¹⁷ *Id.* (quoting James L. Mills & Duane Alexander, *Occasional Notes: Teratogens and "Litogens,"* 315 NEW ENG. J. MED. 1234, 1235 (1986)).

¹¹⁸ *Ferebee* and *Wells* both were decided under the reasonable medical certainty standard. The appellate court's holding in *Elan v. Alcolac* also surprised the scientific community. 765 S.W.2d 42, 212-14 (Mo. Ct. App. 1988). In this case dubious evidence of a controversial clinical ecologist was admitted under the *Frye* test, resulting in a jury verdict of nearly \$50,000,000. *Id.* at 200. Respected scientists greeted the decision with disbelief. Dr. Stuart F. Schlossman, for example, wrote that "the expert testimony in *Alcolac* was not only outside the mainstream of science, it was outside its widest perimeter." Stuart F. Schlossman, *Immunological Tests: A Critique of the Alcolac Decision*, TOXICS LAW REP. 381, 387 (Sept. 6, 1989).

¹¹⁹ *Turpin v. Merrill Dow Pharmaceuticals, Inc.*, 959 F.2d 1349, 1350 (6th Cir. 1982).

¹²⁰ 113 S. Ct. at 2795.

¹²¹ *Id.* Equally surprising outcomes were reached during the course of the Bendectin litigation. In *Oxendine v. Merrell Dow Pharmaceuticals, Inc.*, 506 A.2d 1100 (D.C. 1986), for example, the court made no attempt to scrutinize the validity of the plaintiff's evidence, but relied instead on *Ferebee*. The same outcome was reached by a federal district court judge in *Richardson v. Richardson-Merrell, Inc.*, 649 F. Supp. 799 (D.C. Cir. 1986). These inconsistent results impose a cost. The litigation resulted in Bendectin being removed from the market and may well deter manufacturers from developing new drugs. Black, *supra* note 7, at 681.

A. *The Movement Toward Active Review*

Daubert represents the culmination of a trend in the courts toward active judicial review of the admissibility and sufficiency of scientific evidence, particularly in toxic torts cases. It will accelerate the trend toward active review by mandating an inquiry into the scientific validity of evidence and judicial intervention to prevent a "battle of the experts." Courts in the future will be less likely to rely solely on prior legal precedent or on a scientific head count to determine the admissibility of scientific evidence.

In general, the new standard will lower the threshold for admission of scientific evidence. A lower threshold is prudent because an inquiry into the validity and reliability of scientific evidence requires a fully developed record. Although the new standard will make it easier for plaintiffs to introduce scientific evidence of causation in toxic tort cases, this will prove to be a questionable victory for them. Because the scientific validity standard requires an early active review of the validity and relevance of scientific evidence, it is likely that summary judgment and directed verdict decisions will be used more frequently in toxic tort cases to the detriment of plaintiffs.

There has been a growing trend toward active judicial review of scientific evidence that seems to have been legitimized by the Supreme Court in *Daubert*. This trend seems to have its genesis in Judge Jack Weinstein's landmark decision in *Agent Orange Product Liability Litigation*.¹²² The *Agent Orange* plaintiffs attributed a wide variety of diseases to exposure to dioxin, a contaminant found in Agent Orange, a herbicide used extensively as a defoliant during the Vietnam War.¹²³ Judge Weinstein granted defendants' motion for summary judgment against the claims of several hundred plaintiffs who had opted out of the class settlement.¹²⁴ Judge Weinstein was influenced by the epidemiological evidence presented by the defendants and considered it "the only useful studies having any bearing on causation."¹²⁵ He refused to admit the plaintiffs' countervailing evidence that consisted of animal studies and studies of victims of industrial accidents exposed to high doses.¹²⁶ In particular, he found the animal studies to be of negligible probative force and misleading due to their dubious assumption that an effect found in one species could be extrapolated to humans.¹²⁷

Although many commentators have interpreted the Agent Orange and Bendectin cases as having established judicial activism in preventing the admission of scientific evidence, the more recent trend seems to be dismissal of cases based on the lack of sufficient causation evidence. The Bendectin cases are particularly illustrative of the growing trend toward a discerning review of the sufficiency of scientific evidence. In *Richardson v. Richardson-Merrell, Inc.*, an early Bendectin case, Judge Jackson denied the defendant's motion for summary judgment and allowed both sides to introduce causation evidence.¹²⁸ After hearing the evidence, however, Judge Jackson granted the defendant's motion for judgment notwithstanding the verdict.¹²⁹ The judge was swayed by the extensive body of twenty-one epidemiological studies conducted on Bendectin

¹²² 611 F. Supp. 1223 (E.D.N.Y. 1985), *aff'd on other grounds*, 818 F.2d 187 (2d Cir. 1987), *cert. denied*, 487 U.S. 1234 (1988) (an assessment of this case can be found in Green, *supra* note 15, at 643).

¹²³ Green, *supra* note 15, at 659.

¹²⁴ *Agent Orange*, 611 F. Supp. at 1264.

¹²⁵ *Id.* at 1231.

¹²⁶ *Id.*

¹²⁷ *Id.* at 1241.

¹²⁸ 649 F. Supp. at 800.

¹²⁹ *Id.* at 804.

that found “no statistically significant correlation . . . between pre-natal exposure to Bendectin and the limb reduction defects it was suspected of causing.”¹³⁰ Judge Jackson cited Judge Weinstein’s opinion in *Agent Orange* to bolster his reliance on the epidemiological evidence.¹³¹ He discounted the plaintiff’s re-analysis of extant epidemiological studies, noting that the results had not been subjected to peer review, which tends to reveal methodological errors.¹³² The importance of peer review as an indicator of the validity of scientific evidence was later emphasized by the Supreme Court in *Daubert*.

In light of the large body of human epidemiological evidence, the judge also found that the plaintiff’s proffered animal studies were of questionable value.¹³³ The judge stated that such animal studies are based on the unproven assumption that substances that cause disease in animals are likely to have a similar effect on humans.¹³⁴ Judge Jackson’s view, “the totality of the published scientific literature” on Bendectin “collectively represents all that can be said to be scientifically ‘known’” about its propensity to cause limb reduction birth defects.¹³⁵ He concluded that “there is now nearly universal scientific consensus that Bendectin has not been shown to be a teratogen, and, the issue being a scientific one, reasonable jurors could not reject that consensus without indulging in [speculation].”¹³⁶

In a subsequent Bendectin case, *Lynch v. Merrell-national Laboratories Division of Richardson-Merrell, Inc.*,¹³⁷ the First Circuit also engaged in active review and arguably went too far in rejecting all non-epidemiological evidence.¹³⁸ While the *Richardson* trial court dismissed the case because the plaintiff did not introduce sufficient evidence to prove causation, the *Lynch* court refused to admit most of the plaintiff’s causation evidence, instead focusing on the extensive epidemiological evidence offered by the parties. The court was impressed by the strength of the epidemiological evidence, finding that Bendectin was not a teratogen and upholding the district court’s grant of defendant’s motion for summary judgment.¹³⁹ The court carefully reviewed the scientific studies and the evidence offered, and concluded that the “association of Bendectin with limb reduction” is “an instance of popular delusion and error.”¹⁴⁰

The extensive body of epidemiological evidence that was available to the *Lynch* court is rare in a toxic tort case. The animal studies and chemical structure/activity analysis the plaintiff sought to introduce assume more importance when the epidemiological evidence is undeveloped. The *Lynch* court’s conclusion that “studies of this sort singly or in combination, do not have the capability of proving causation in human

¹³⁰ *Id.* at 802.

¹³¹ *Id.*

¹³² *Id.*

¹³³ *Id.*

¹³⁴ Other problems with animal studies are pointed out by Susan Poulter: “Of 165 substances studied with no teratogenic finding in humans, only 47 or 29%, were negative in all laboratory animal test species.” Susan R. Poulter, *Science and Toxic Torts: Is There a Rational Solution to the Problem of Causation?*, 7 HIGH TECH. L.J. 189, 221 n.154 (1992). Poulter also points out that animal studies conducted to identify carcinogenicity “are conducted under conditions very different from the usual human exposure scenario. Animal studies . . . typically utilize doses at or near the maximum level tolerated by the animal . . . [and are conducted on] animal strains bred for their susceptibility for tumor formation.” *Id.* at 203.

¹³⁵ 649 F. Supp. at 802.

¹³⁶ *Id.* at 803.

¹³⁷ 830 F.2d 1190 (1st Cir. 1987).

¹³⁸ Green, *supra* note 15, at 665.

¹³⁹ 830 F.2d at 1193.

¹⁴⁰ *Id.* at 1194.

beings in the absence of confirmatory epidemiological data” is overbroad.¹⁴¹ These forms of evidence serve as a valuable supplement to the more modest epidemiological evidence available in the typical toxic tort case. The court’s rejection of this evidence may have been appropriate in this particular case, but is inappropriate for the majority of toxic tort cases.

The most far-reaching decision in the Bendectin line of cases is *Brock v. Merrell Dow Pharmaceuticals, Inc.*, in which the Fifth Circuit reversed a jury verdict in favor of the plaintiff.¹⁴² The court did not conduct a thorough review of the defendant’s epidemiological evidence as occurred in *Richardson* and *Lynch*, but rather focused on weaknesses in the plaintiff’s evidence.¹⁴³ The court erroneously concluded that the concept of a confidence interval accounted for all sources of error in an epidemiological study when only random sampling errors are incorporated.¹⁴⁴ Specifically, the confidence interval does not account for systematic errors such as recall bias and the existence of confounding factors as suggested by the *Brock* court.¹⁴⁵ Relying on its mistaken understanding of epidemiological evidence, the court concluded that “the lack of conclusive epidemiological proof was fatal to the Brock’s case.”¹⁴⁶

Green interprets the *Brock* court as mandating a “statistically significant epidemiological threshold” as a precondition to submitting the causation issue to the jury.¹⁴⁷ The *Brock* court may not have intended such a reading, as it also remarks that epidemiological proof is not a necessary element in a successful toxic tort suit.¹⁴⁸ Green notes that the Bendectin cases are unusual in toxic tort litigation in that birth defects have an unusually short latency period.¹⁴⁹ The longer latency period of diseases such as cancer, combined with the high cost of epidemiological studies, operates to ensure that the epidemiological record rarely will be as well developed as in the Bendectin context.¹⁵⁰ Alternate forms of evidence should not be dismissed when the epidemiological record is underdeveloped, as is the norm. Plaintiffs should be permitted to “prove causation by the preponderance of the available evidence.”¹⁵¹ In *Daubert* the Supreme Court did not follow the lead of *Brock* and did not insist on epidemiological evidence as a precondition for proving causation.

The Sixth Circuit’s decision in *Turpin v. Merrell Dow* is the most recent addition to the Bendectin line of cases.¹⁵² The decision is significant because the Supreme Court cited it with approval in *Daubert*. In *Turpin*, the court affirmed the district court’s grant of a motion for summary judgment on the grounds that the plaintiff’s causation evidence, particularly animal studies, “[were] insufficient to allow a rational jury to find that Bendectin caused the minor plaintiffs’ birth defects.”¹⁵³ The court concluded that it had a duty to take a “hard look” at the scientific evidence offered in the case “believ[ing] that close judicial analysis of such technical and specialized matter is necessary not

¹⁴¹ *Id.*

¹⁴² 874 F.2d at 308.

¹⁴³ *Id.* at 312-13.

¹⁴⁴ *Id.* at 312.

¹⁴⁵ *Id.*

¹⁴⁶ *Id.* at 313.

¹⁴⁷ Green, *supra* note 15, at 672.

¹⁴⁸ 874 F.2d at 313.

¹⁴⁹ Green, *supra* note 15, at 679.

¹⁵⁰ *Id.* at 679 (quoting JOSEPH GASTWIRTH, STATISTICAL REASONING IN LAW AND PUBLIC POLICY 839 (1988)).

¹⁵¹ *Id.* at 680.

¹⁵² 959 F.2d at 1349.

¹⁵³ *Id.* at 1350.

only because of the likelihood of juror misunderstanding, but also because expert witnesses are not necessarily always unbiased scientists."¹⁵⁴ In language similar to that later used by the Supreme Court, the court stated that "courts have a duty to inspect the reasoning of qualified scientific experts to determine whether a case should go to the jury."¹⁵⁵ The court began its inquiry into the scientific validity and reliability of causation evidence by reviewing the results of several of the thirty-five epidemiological studies conducted on the effects of Bendectin.¹⁵⁶ The court concluded that although the studies failed to find a statistically significant association between Bendectin and birth defects, it was still possible that Bendectin was a teratogen.¹⁵⁷ The court noted that the incidence of limb reduction defects was so rare that a large number of births would have to be considered to eliminate Bendectin as a teratogen.¹⁵⁸ The court then turned its attention to the *in vitro* and *in vivo* animal studies offered by the plaintiff. The court rejected the *in vivo* studies because "different species of animals react differently to the same stimuli for reasons not entirely understood."¹⁵⁹ The court found the testimony of plaintiffs' experts based on *in vitro* studies inadequate because the expert testified only to possibility and "not that they do cause such defects."¹⁶⁰ The court finally concluded that the "analytical gap between the evidence presented and the ultimate issue of human birth defects [was] too wide" and dismissed the case for insufficient proof of causation.¹⁶¹

The most recent Bendectin cases, including *Turpin* and *Brock*, exemplify a trend toward dismissal of cases based on lack of sufficient evidence. Additionally, the *Brock*, *Lynch*, and *Richardson* line of analysis foreshadows the Supreme Court's reasoning in *Daubert* in several respects. The *Brock* court, for example, determined that "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."¹⁶² The focus in *Brock*, as in *Daubert*, is on scientific validity and reliability. The courts in *Brock*, *Lynch*, and *Richardson* were skeptical of re-analysis of epidemiological and other types of evidence that had not been published and subjected to peer review. The Supreme Court similarly showed respect for the peer review process by listing peer review as one of the four indicia of reliability in the *Daubert* opinion.

B. *Daubert* Continues the Bendectin Trend Toward Active Judicial Review of the Sufficiency of Evidence

Examining the earliest cases that have utilized the Supreme Court's holding in *Daubert*, it appears that the decision will be used to rule on the sufficiency, as often as the admissibility, of evidence. In this respect, courts have followed the trend evident in the most recent Bendectin cases, including *Turpin* and *Brock*, toward dismissal based on lack of sufficient causation evidence.

Two of the early courts to apply the *Daubert* decision in a toxic tort context have

¹⁵⁴ *Id.* at 1352.

¹⁵⁵ *Id.*

¹⁵⁶ *Id.* at 1356.

¹⁵⁷ *Id.* at 1357.

¹⁵⁸ *Id.*

¹⁵⁹ *Id.* at 1359.

¹⁶⁰ *Id.* at 1358.

¹⁶¹ *Id.* at 1360.

¹⁶² 874 F.2d at 310 n.7.

read dicta in the case as promoting stringent judicial review of the sufficiency of the plaintiff's proffered evidence. In the first of these two cases, *In re Joint Eastern and Southern District Asbestos Litigation*, Judge Sweet conceded that "there is no question in this case that the testimony of both side's experts . . . is admissible."¹⁶³ The judge then interpreted dictum from the *Daubert* opinion as requiring the court to make its "own assessment of whether the methodology underlying the expert's opinion is fundamentally sound" in order to determine the *sufficiency* of the evidence.¹⁶⁴ The judge seemingly viewed the *Daubert* opinion as requiring it to exercise vigilance in screening junk science. According to the district court, the Supreme Court "[did] address the question of the sufficiency of the evidence" in its *Daubert* opinion:

[If] the trial court concludes that the scintilla of evidence presented supporting a position is insufficient to allow a reasonable juror to conclude that the position more likely than not is true, the court remains free to direct a judgment. These conventional devices, rather than wholesale exclusion under an uncompromising "general acceptance" test, are the appropriate safeguards.¹⁶⁵

The district court also relied on the fact that the Supreme Court cited *Turpin* and *DeLuca v. Merrell Dow Pharmaceuticals, Inc.*¹⁶⁶ approvingly as sound decisions in which courts conducted probing reviews of the sufficiency of the evidence prior to granting judgments as a matter of law.¹⁶⁷ Judge Sweet concluded from these precedents that "the epidemiological, clinical, and experimental studies testified to at trial by the plaintiff's experts must be carefully scrutinized to determine whether they are, in fact, sufficient."¹⁶⁸ He then analyzed the plaintiff's evidence using the sufficiency criteria commonly employed by epidemiologists and found the evidence insufficient to carry the plaintiff's burden of proof with respect to causation.¹⁶⁹

The United States Claims Court, Office of Special Matters, while not bound by the Federal Rules of Evidence, also seized on the Supreme Court's decision in *Daubert* to determine the sufficiency of the plaintiff's causation evidence in a toxic tort case. In *Haim v. Secretary of the Department of Health and Human Services*,¹⁷⁰ the plaintiff sued for compensation under the National Childhood Vaccine Injury Act of 1986,¹⁷¹ alleging that the DPT vaccine administered to her infant caused an encephalopathy, which ultimately resulted in the child's death.¹⁷² The primary evidence relied on by the plaintiff was the testimony of two physicians who examined the child's extensive medical record.¹⁷³ The testimony of both physicians relied heavily on the National Childhood Encephalopathy Study (NCES),¹⁷⁴ which suggests that the DPT vaccine can cause

¹⁶³ 827 F. Supp. at 1025.

¹⁶⁴ *Id.* at 1031.

¹⁶⁵ *Id.* at 1033 (quoting *Daubert*, 113 S. Ct. at 2798).

¹⁶⁶ 911 F.2d 941 (3d Cir. 1990).

¹⁶⁷ 827 F. Supp. at 1033.

¹⁶⁸ *Id.* at 1037.

¹⁶⁹ See *supra* notes 44-59 and accompanying text for a description of the sufficiency criteria and their application in the present case.

¹⁷⁰ 90-1031V, WL 346392 (Cl. Ct. Aug. 27, 1993).

¹⁷¹ Pub. L. No. 99-660, 100 Stat. 3743 (codified at 42 U.S.C. §§ 300aa et seq. (1988)).

¹⁷² WL 346392 at *1.

¹⁷³ *Id.* at *3.

¹⁷⁴ ROBERT ALDERSLADE, UNITED KINGDOM DEPARTMENT OF HEALTH AND SOCIAL SECURITY, WHOOPING COUGH REPORT 79-169 (1981).

neurological problems.¹⁷⁵ Dr. Geir, one of the plaintiff's experts, also based his testimony on animal evidence.¹⁷⁶

The court conceded that the issue in *Daubert* "concerned the admissibility of scientific evidence, whereas the present focus of this court is in weighing previously admitted evidence."¹⁷⁷ Nevertheless the court determined that *Daubert* "provides cogent guidelines regarding the validity of scientific evidence" and applied these guidelines to the sufficiency of the plaintiff's causation evidence.¹⁷⁸ The court held that the NCES and animal studies were flawed, rendering invalid any opinion resting on them.¹⁷⁹ The NCES study suffered from several methodological problems noted by the court and its authors.¹⁸⁰ In addition, it was statistically insignificant.¹⁸¹ In view of the weakness of the underlying studies, the court found that Dr. Geir's testimony did not constitute scientific knowledge under *Daubert*, but rather was "merely subjective belief and unsupported speculation."¹⁸² The *Haim* opinion is a striking illustration of how far the role of a judge in scrutinizing scientific evidence has departed from the traditional deference afforded by the reasonable medical certainty standard as reflected in *Wells* and *Ferebee*.

V. THE DAUBERT DECISION WILL EXACERBATE THE FIRST PLAINTIFF PROBLEM

The first plaintiff bears a heavy burden in toxic tort litigation as he or she must find legal counsel willing to take the case, often for a contingency fee, and must locate scarce scientific evidence demonstrating causation between his or her injuries and the suspect substance.¹⁸³ At this stage, defendants possess substantial advantages because they usually have superior financial resources and access to the best available scientific evidence.¹⁸⁴ The defendant manufacturer or pharmaceutical company often has generated the majority of the available evidence in seeking government agency approval of the product.¹⁸⁵ The Supreme Court's decision in *Daubert* will contribute to the plaintiff's disadvantages at the outset of litigation by increasing the likelihood of active judicial review to exclude the plaintiff's weak causation evidence. Prior to the new standard, the courts occasionally would allow weak or questionable scientific evidence to reach the jury, resulting in an early victory for plaintiff's that then was relied on by following plaintiffs to avoid summary judgment, attract competent counsel, and increase the settlement value of a case. An early victory of this type occurred in the Bendectin cases when the court in *Oxendine v. Merrell Dow Pharmaceuticals, Inc.*¹⁸⁶ sustained a jury verdict based on weak causation evidence. *Daubert* shifts responsibility for determining the validity of scientific evidence away from the jury to the judge, mitigating against plaintiff victories of this kind. Ironically, the new standard, while increasing the amount of

¹⁷⁵ *Id.* at 143.

¹⁷⁶ WL 346392 at *10.

¹⁷⁷ *Id.*

¹⁷⁸ *Id.* at *10-15.

¹⁷⁹ *Id.* at *14.

¹⁸⁰ *Id.* at *10-14.

¹⁸¹ ALDERSLADE, *supra* note 174, at 141.

¹⁸² WL 346392 at *15.

¹⁸³ Joseph Sanders, *The Bendectin Litigation: A Case Study in the Life Cycle of Mass Torts*, 43 HASTINGS L.J. 301, 349 (1992).

¹⁸⁴ *Id.* at 350.

¹⁸⁵ *Id.*

¹⁸⁶ 506 A.2d at 1100.

evidence that will be admitted, portends an increased willingness of courts to grant summary judgment and judgment notwithstanding the verdict as evidenced in the early post-*Daubert* decisions.

VI. CONCLUSION

The Supreme Court's opinion in *Daubert v. Merrell Dow Pharmaceuticals, Inc.* replaces the long-standing *Frye* test governing the admissibility of scientific evidence with a new standard that places a premium on scientific validity and relevance. The *Frye* standard proved to be vague in several respects. First, the test did not define criteria to identify the pertinent scientific community or the level of agreement needed to establish "general acceptance in a particular field." Second, the test was applied only to novel techniques. This limitation permitted courts to avoid the rigors of the *Frye* standard. The inherent vagueness and malleability of the general acceptance test resulted in outcomes that often conflict with the consensus of the scientific community.

The *Daubert* standard represents a significant improvement over the *Frye* and reasonable medical certainty tests previously used to scrutinize expert testimony in toxic tort cases. The standard correctly focuses judicial attention on the real issue: whether the proffered testimony is based on scientifically valid reasoning connecting facts to conclusions. The inquiry required by the test will force judges to scrutinize scientific studies and evidence with which they are uncomfortable. This is preferable to allowing a case to degenerate into a battle of experts with juries reaching conclusions that purport to rely on science, but which are based on questionable reasoning and are at odds with the views of the scientific community as expressed in literature. The deficiencies of the *Frye* and reasonable medical certainty standards are starkly apparent in two recent cases, *Ferebee v. Chevron Chemical* and *Wells v. Ortho Pharmaceutical Corp.* These cases skirted the issue of scientific validity and left juries to speculate and reach conclusions that did not comport with the state of scientific knowledge. Once fully understood by lower courts, the *Daubert* standard will provide an effective means to screen evidence for conformance to scientific method and accepted scientific practice. The *Daubert* standard should result in the admission of only scientifically sound evidence, leading to decisions consistent with science and with each other.

Consistent with the thrust of the Federal Rules of Evidence emphasized by the Court in establishing the new standard, the *Daubert* test will lower the threshold for admission of scientific evidence. A lower threshold is prudent because an inquiry into the validity and reliability of scientific evidence requires a fully-developed record. Although the new standard will make it easier for plaintiffs to introduce scientific evidence of causation in toxic tort cases, this ironically will prove to be a questionable victory for plaintiffs. Because the scientific validity standard calls for early active review of the validity and relevance of scientific evidence, it is likely that summary judgment and directed verdicts will be granted more frequently in toxic tort cases, to the detriment of plaintiffs.

Beyond the Court's focus on scientific validity in addressing the admissibility of scientific evidence, the *Daubert* standard is a flexible one that must be tailored to fit the type of evidence under consideration. The cases decided to date under the new standard demonstrate that this challenge is within the competence of courts.