



2001

The Dynamics of Daubert: Methodology, Conclusions, and Fit in Statistical and Econometric Studies

David H. Kaye
Penn State Law

Follow this and additional works at: http://elibrary.law.psu.edu/fac_works

 Part of the [Antitrust and Trade Regulation Commons](#), [Evidence Commons](#), and the [Science and Technology Law Commons](#)

Recommended Citation

David H. Kaye, *The Dynamics of Daubert: Methodology, Conclusions, and Fit in Statistical and Econometric Studies*, 87 *Va. L. Rev.* 1933 (2001).

This Article is brought to you for free and open access by the Faculty Works at Penn State Law eLibrary. It has been accepted for inclusion in Journal Articles by an authorized administrator of Penn State Law eLibrary. For more information, please contact ram6023@psu.edu.

THE DYNAMICS OF *DAUBERT*:
 METHODOLOGY, CONCLUSIONS, AND FIT IN
 STATISTICAL AND ECONOMETRIC STUDIES

*D.H. Kaye**

INTRODUCTION.....	1934
I. STANDARDS FOR ADMITTING SCIENTIFIC EVIDENCE.....	1937
<i>A. The Classical Period: Relevant Expertise</i>	<i>1938</i>
<i>B. The Modern Period: Heightened Scrutiny for Scientific Evidence.....</i>	<i>1943</i>
1. <i>General Acceptance: Frye.....</i>	<i>1945</i>
2. <i>Relevancy-Plus: The Road to Daubert.....</i>	<i>1956</i>
3. <i>Scientific Soundness: Daubert.....</i>	<i>1958</i>
<i>C. The Puzzles of Strict Scrutiny.....</i>	<i>1964</i>
1. <i>The Boundary Problem.....</i>	<i>1964</i>
2. <i>The Usurpation Problem and the Methodology- Conclusion Puzzle.....</i>	<i>1972</i>
<i>D. Looking Back at Statistical Evidence</i>	<i>1985</i>
II. BRINGING <i>DAUBERT</i> AND <i>KUMHO</i> TO BEAR: <i>CONWOOD CO.</i>	
<i>V. UNITED STATES TOBACCO CO.....</i>	<i>1988</i>
<i>A. Conwood's Complaint: Monopolizing Moist Snuff.....</i>	<i>1988</i>
<i>B. Conwood's Resistance Theory.....</i>	<i>1989</i>
<i>C. Conwood's Data.....</i>	<i>1990</i>
<i>D. Regression Analysis to Show Causation.....</i>	<i>1992</i>
1. <i>The Regression Results.....</i>	<i>1992</i>
2. <i>The Causal Inference.....</i>	<i>1996</i>
<i>E. Regression Analysis to Estimate Damages.....</i>	<i>2002</i>
1. <i>Estimating Effect with a "Regression Rectangle"</i>	<i>2002</i>
2. <i>Applying Daubert to the "Regression Rectangle"</i>	<i>2006</i>
3. <i>"Internal" Criticisms of the Regression</i>	<i>2011</i>
CONCLUSION.....	2013
APPENDIX.....	2015

* Regents' Professor, College of Law, Arizona State University; Fellow, Center for the Study of Law, Science, and Technology. I am grateful to David Freedman for comments on a draft of this paper.

INTRODUCTION

IN *Daubert v. Merrell Dow Pharmaceuticals, Inc.*,¹ the Supreme Court stated the obvious—trial judges have a “gatekeeping role”² when it comes to scientific evidence. The Court’s conclusion—that the Federal Rules of Evidence dispense with the “general acceptance” standard that previously dominated the field—is less obvious.³ Still, the “reliability” standard announced in *Daubert* was nothing new. Rather, this standard reiterates the law as it then stood in many jurisdictions.⁴ The striking feature of both the reliability and the general acceptance standards is that the court must subject “scientific” evidence to heightened scrutiny.⁵ This approach creates two broad problems: (1) the “boundary problem” of identifying the type of evidence that warrants such careful screening;⁶ and (2) the “usurpation problem” of keeping the trial judge from closing the gate on evidence that should be left for the jury to assess.⁷

Being less revolutionary than one might think from the volumes that have been written about it, *Daubert* does little to resolve these perdurable puzzles and problems. The Supreme Court’s more recent opinion in *Kumho Tire Co. v. Carmichael*⁸ sidesteps the boundary problem by making the reliability standard applicable to all expert testimony⁹ and demanding more “rigor” for all expert

¹ 509 U.S. 579 (1993).

² *Id.* at 597.

³ See *infra* Section I.B.3.

⁴ See *infra* Section I.B.2.

⁵ “Strict scrutiny” or “heightened scrutiny” are phrases originally developed in the field of constitutional law to describe the level of judicial review of legislation. Here, I use the terms to denote an unusually demanding level of review for the admissibility of scientific evidence as opposed to expert testimony in general.

⁶ D.H. Kaye, Choice and Boundary Problems in *Logerquist, Hummert, and Kumho Tire*, 33 *Ariz. St. L.J.* 41 (2001).

⁷ See *Logerquist v. McVey*, 1 P.3d 113, 131 (Ariz. 2000) (criticizing *Daubert* and its extension to technical and other forms of expertise on the ground that “[t]he right to jury trial does not turn on the judge’s preliminary assessment of testimonial reliability. It is the jury’s function to determine accuracy, weight, or credibility.”); Edward J. Imwinkelried, *Logerquist v. McVey: The Majority’s Flawed Procedural Assumptions*, 33 *Ariz. St. L.J.* 121 (2001) (arguing that Federal Rule of Evidence 104 solves the problem).

⁸ 526 U.S. 137 (1999).

⁹ Exactly what this means is considered *infra* Section I.C.2.

testimony.¹⁰ The emphasis on intellectual rigor, however, has the potential to exacerbate the usurpation problem.¹¹ The threat is intensified by the Court's opinion in *General Electric Co. v. Joiner*,¹² which encourages the trial court to exclude testimony when it disagrees with the expert's conclusions as well as the underlying scientific method.¹³

This paper will propose at least partial solutions to the boundary and usurpation problems, and it will apply them to statistical and econometric proof. In addition, the paper will review the developments that have culminated in the modern use of sophisticated statistical equations and models to prove factual claims such as the presence of illegal discrimination,¹⁴ racial polarization in voting,¹⁵ the identity of criminals,¹⁶ the existence of forgeries,¹⁷ the causes of

¹⁰ *Kumho*, 526 U.S. at 152.

¹¹ See *Logerquist*, 1 P.3d at 125 ("The result reached in *Kumho*, however, would seem directly opposed to the principle of liberalized admissibility that engendered the abolition of *Frye*.").

¹² 522 U.S. 136 (1997).

¹³ See *infra* Section I.C.2.

¹⁴ See, e.g., *Statistical Methods in Discrimination Litigation* 69–84 (D.H. Kaye & Mikel Aickin eds., 1986); Richard Lempert, *Befuddled Judges: Statistical Evidence in Title VII Cases*, in *Legacies of the 1964 Civil Rights Act* 263 (Bernard Grofman ed., 2000).

¹⁵ See, e.g., *Thornburg v. Gingles*, 478 U.S. 30 (1986); Alan D. Taylor, *Mathematics and Politics: Strategy, Voting, Power, and Proof* (1995); Bernard Grofman, *Expert Witness Testimony and the Evolution of Voting Rights Case Law*, in *Controversies in Minority Voting: The Voting Rights Act in Perspective* 197 (Bernard Grofman & Chandler Davidson eds., 1992); David H. Kaye & David A. Freedman, *Reference Guide on Statistics*, in *Reference Manual on Scientific Evidence* 83, 141–43 (2d ed. 2000).

¹⁶ See, e.g., National Research Council, *The Evaluation of Forensic DNA Evidence* (1996) [hereinafter *NRC Report*] (focusing on population genetics and statistical issues, with recommendations); D.H. Kaye, *Science in Evidence* 195–221 (1997); Hans Zeisel & David Kaye, *Prove It with Figures: Empirical Methods in Law and Litigation* 199–224 (1997) (discussing DNA profiling in criminal cases); cf. D.H. Kaye, *The Admissibility of "Probability Evidence" in Criminal Trials* (pts. 1 & 2), 26 *Jurimetrics J.* 343 (1986), 27 *Jurimetrics J.* 160 (1987) (discussing other, often less sophisticated uses of probability theory to link a defendant to a crime); D.H. Kaye, *Statistics for Lawyers and Law for Statistics*, 89 *Mich. L. Rev.* 1520, 1525–44 (1991) (discussing an attempt to show mathematically that the defendant could not have traveled from work to home and murdered his wife in the time period allowed by the state's evidence).

¹⁷ See, e.g., Paul Meier & Sandy Zabell, *Benjamin Peirce and the Howland Will*, 75 *J. Am. Stat. Ass'n* 497 (1980); *The Howland Will Case*, 4 *Am. L. Rev.* 625 (1870).

trends in sales or prices,¹⁸ and the quantum of damages caused by illegal conduct.¹⁹

Part I will show that, before *Daubert*, the admissibility of complex statistical evidence usually was taken for granted. Arguments centered on the weight to be accorded to this evidence in particular cases. Today, pretrial motions challenging the admissibility of statistical studies have become commonplace. Federal courts are now faced with the task of fitting this type of expertise into the framework for determining admissibility constructed in the *Daubert-Joiner-Kumho* trilogy and codified in the Federal Rules of Evidence.²⁰ Part I will also analyze the admissibility issue under other standards for screening scientific evidence. Some states that use a scientific validity standard à la *Daubert* might not follow *Kumho*. These jurisdictions will have to determine whether statistics and economics are subject to any form of heightened scrutiny. Some states might resist *Joiner's* blurring of the distinction between methodology and conclusion. These jurisdictions will have to decide what aspects of statistical testimony constitute the methodology that must be scientifically valid. Finally, states that adhere to the standard of general scientific acceptance face compa-

¹⁸ See, e.g., Daniel L. Rubinfeld & Peter O. Steiner, Quantitative Methods in Antitrust Litigation, 46 Law & Contemp. Probs. 69 (1983).

¹⁹ See, e.g., *Spray-Rite Serv. Corp. v. Monsanto Co.*, 684 F.2d 1226, 1240-41 (8th Cir. 1982); *Litigation Economics* (Patrick A. Gaughan & Robert J. Thornton eds., 1993); R.S. Daggett & D.A. Freedman, *Econometrics and the Law: A Case Study in the Proof of Antitrust Damages*, in 1 Proceedings of the Berkeley Conference in Honor of Jerzey Neyman 123 (Lucien M. Le Cam & Richard A. Olshen eds., 1985); Michael O. Finkelstein & Hans Levenbach, *Regression Estimates of Damages in Price-Fixing Cases*, 46 Law & Contemp. Probs. 145 (1983), reprinted in *Statistics and the Law* 79 (Morris H. DeGroot et al. eds., 1986); Scott L. Zeger et al., *Statistical Testimony on Damages in Minnesota v. Tobacco Industry*, in *Statistical Science in the Courtroom* 303 (Joseph L. Gastwirth ed., 2000).

For general discussions of statistics in litigation or collections of papers on the topic, see, for example, *The Evolving Role of Statistical Assessments as Evidence in the Courts* (Stephen E. Fienberg ed., 1989) [hereinafter *Statistical Assessments as Evidence*]; *Statistical Science in the Courtroom* (Joseph L. Gastwirth ed., 2000); *Statistics and the Law*, supra. Expositions of statistical theory tailored to legal applications include Michael O. Finkelstein & Bruce Levin, *Statistics for Lawyers* (2d ed. 2001); 1 Joseph L. Gastwirth, *Statistical Reasoning in Law and Public Policy* (1988); Zeisel & Kaye, supra note 16, at 45-68; Kaye & Freedman, supra note 15; Daniel L. Rubinfeld, *Reference Guide on Multiple Regression*, in *Reference Manual on Scientific Evidence*, supra note 15, at 179.

²⁰ Fed. R. Evid. 702 (2000).

rable challenges in defining the subject matter of statistics and economics and the scope of this test for the admissibility of scientific evidence.

After analyzing the leading cases on scientific evidence and discussing their effects on efforts to introduce statistical proof, this paper will consider these emerging issues in the context of an anti-trust case in which an econometric analysis was introduced to show both causation and damages. By describing the arguments on a pending appeal, Part II will illustrate the difficulty of distinguishing between statistical methodology and conclusion, but will conclude that the distinction is viable and valuable. The discussion will also reveal the extent to which the dictum in *Kumho* concerning the need for rigor encourages arguments as to admissibility that, in an earlier era, would have been treated as affecting only the weight of expert evidence. Finally, the case shows how difficult it can be to explain to judges and juries serious methodological defects in statistical assessments.

The paper will conclude that *Daubert*-like screening of complex statistical analyses is a salutary development, but that the task requires the elaboration of standards that attend to the distinction between a general methodology and a specific conclusion. Screening statistical proof demands some sophistication in evaluating the choice of a research design or statistical model, the variables included in a particular model, the procedures taken to verify the usefulness of the model for the data at hand, and the inferences or estimates that follow from the statistical analysis. The factors enumerated in *Daubert* work reasonably well with some of these aspects of the expert's work, but these factors are less well adapted to other aspects. If the "intellectual rigor" standard of *Kumho* is used to fill the gap, it must be applied with some caution lest it become a subterfuge for excluding expert testimony that is less than ideal but still within the range of reasonable scientific debate.

I. STANDARDS FOR ADMITTING SCIENTIFIC EVIDENCE

Statistics are part of science, and science is one type of expertise. To appreciate how the law of evidence pertains to statistical proof, we must consider the rules of evidence as they apply to experts in general and to scientific testimony in particular. With that necessary prolegomenon, we will be in a position to determine how

these approaches to regulating scientific and expert testimony have been and should be applied to statistical and econometric proof.

A. The Classical Period: Relevant Expertise

For centuries, the law did not distinguish one type of expert testimony from another.²¹ On the surface, a uniform standard governed the admission of the testimony of all qualified experts.²² The evidence had to be relevant and not too prejudicial or time-consuming, and it had to deal with matters not comprehensible to ordinary jurors without the assistance of an expert. A few jurisdictions continue in this tradition,²³ although the beyond-the-ken-of-

²¹ See, e.g., 3 Simon Greenleaf, *A Treatise on the Law of Evidence* § 440, at 483 (1858) (describing the admissibility of opinion testimony from all "persons of skill, sometimes called *experts*," on "questions of science, skill, or trade, or others of the like kind"). Before the sixteenth century, experts usually were part of the jury. Once experts became witnesses, it is not clear whether they were called by the parties or the court. John Basten, *The Court Expert in Civil Trials—A Comparative Appraisal*, 40 *Mod. L. Rev.* 174, 175–76 (1977). The first clear reference to an expert witness called by and on behalf of a party appears in *Folkes v. Chadd*, 99 *Eng. Rep.* 589, 589–90 (1782). Basten, *supra*, at 176.

²² For early discussions of expert testimony, see Lee M. Friedman, *Expert Testimony, Its Abuses and Reformation*, 19 *Yale L.J.* 247 (1910) (proposing stronger guarantees against the admission of interested or unqualified expert testimony); Learned Hand, *Historical and Practical Considerations Regarding Expert Testimony*, 15 *Harv. L. Rev.* 40 (1901) (discussing the development of the use of expert witnesses); Clemens Herschel, *Services of Experts in the Conduct of Judicial Inquiries*, 21 *Am. L. Rev.* 571 (1887).

²³ E.g., *Dow Chem. Co. v. Mahlum*, 970 P.2d 98, 108 n.3 (Nev. 1998) (declining to adopt *Daubert* even though it is persuasive because of a belief that the doctrine is a "work in progress"); *State v. Council*, 515 S.E.2d 508, 517–18 (S.C. 1999); *State v. Peters*, 534 N.W.2d 867, 872–73 (Wis. Ct. App. 1995) ("Unlike judges in *Frye* and *Daubert* jurisdictions, this role is much more oblique and does not involve a direct determination as to the reliability of the scientific principle on which the evidence is based.").

An issue often ignored under the classical relevancy-expertise standard is whether the putative expertise exists. Emphasizing that most experts were drawn from professions or occupations in which they had achieved "a modicum of prosperity," however, several commentators have maintained that an implicit "commercial marketplace test" was the usual means of demonstrating the existence of a body of expert knowledge or skill. 1 *Modern Scientific Evidence: The Law and Science of Expert Testimony* § 1-2.1, at 3 & n.5 (David L. Faigman et al. eds., 1997) [hereinafter *Modern Scientific Evidence*]. Indeed, this treatise suggests that the commercial marketplace was the sole source for recognized expertise. *Id.* at 3–4 (drawing on David L. Faigman et al., *Check Your Crystal Ball at the Courthouse Door, Please: Exploring the Past, Understanding the Present, and Worrying about the Future of*

the-jury standard²⁴ usually has been softened to require only that the expert's knowledge be helpful to the jury.²⁵

Although the relevance-expertise requirement applies to scientific and nonscientific expertise alike, it need not have the same impact on all types of expert testimony. Scientific evidence tends to be time-consuming and difficult to understand.²⁶ Courts fear that it comes cloaked in an aura of infallibility and that this leads jurors to give it more credence than it deserves.²⁷ Consequently, ad hoc bal-

Scientific Evidence, 15 *Cardozo L. Rev.* 1799 (1994)). The authorities cited for this strong claim, however, are inconclusive. See *id.*, §1-2.1, at 4 n.7. Furthermore, there are indications that courts were quite willing to recognize expertise acquired outside the commercial marketplace. See, e.g., 3 *Greenleaf*, *supra* note 21, at 485 (referring to “[a] person acquainted for many years with a certain stream, its rapidity of rise in times of freshet, and the volume and force of its waters” as entitled to give an expert opinion in “the sufficiency of a dam, erected in that place, to resist the force of the flood”).

²⁴ E.g., *State v. Kelly*, 478 A.2d 364, 379 (N.J. 1984) (“[T]he intended testimony must concern a subject matter that is beyond the ken of the average juror.”); 3 *Greenleaf*, *supra* note 21, § 440a, at 487 (“The testimony of experts is not admissible upon matters of judgment within the knowledge and experience of ordinary jurymen”); Charles T. McCormick, *Handbook of the Law of Evidence* § 13, at 28 (1954) (“[T]he subject . . . must be . . . beyond the ken of the average layman.”).

²⁵ See, e.g., 1 *McCormick on Evidence* § 13, at 58–59 (John W. Strong ed., 5th ed. 1999).

²⁶ See, e.g., *United States v. Hall*, 165 F.3d 1095, 1119 (7th Cir. 1999) (Easterbrook, J., concurring) (“Social science evidence is difficult to absorb; the idea of hypothesis formulation and testing is alien to most persons. That’s one reason why the training of social scientists is so extended.”).

²⁷ See, e.g., *United States v. Addison*, 498 F.2d 741, 744 (D.C. Cir. 1974) (stating that scientific evidence may “assume a posture of mythic infallibility in the eyes of a jury of laymen”); *United States v. Amaral*, 488 F.2d 1148, 1152 (9th Cir. 1973) (noting the potential prejudicial effect arising from the “aura of special reliability and trustworthiness” of scientific testimony); Christopher B. Mueller & Laird C. Kirkpatrick, *Modern Evidence: Doctrine and Practice* § 7.8, at 992 (1995) (citing cases involving efforts to procure funds for expert testimony on eyewitness identification for the proposition that “[s]cientific proof may suggest unwarranted certainty to lay factfinders, especially if it comes dressed up in technical jargon, complicated mathematical or statistical analysis, or involves a magic machine (‘black box’) that may seem to promise more than it delivers”); John William Strong, *Language and Logic in Expert Testimony: Limiting Expert Testimony by Restrictions of Function, Reliability, and Form*, 71 *Or. L. Rev.* 349, 367 n.81 (1992) (“There is virtual unanimity among courts and commentators that evidence perceived by jurors to be ‘scientific’ in nature will have particularly persuasive effect.”). Some commentators, skeptical of the claims of mainstream science and medicine that these disciplines know the truth, have more faith in the power of juries to evaluate contested scientific evidence. See, e.g., Michael S. Jacobs, *Testing the Assumptions Underlying the Debate About Scientific Evidence: A Closer Look at Juror “Incompetence” and Scientific “Objectivity,”* 25 *Conn. L. Rev.* 1083 (1993).

ancing of probative value and its counterweights can operate to exclude scientific evidence, especially if the science is not well-established.²⁸

Perhaps the earliest reported instance of a statistical assessment admitted under this classical approach is *Robinson v. Mandell*.²⁹ On July 2, 1865, Sylvia Ann Howland died. An 1863 will left half the estate, worth more than two million dollars, to a number of individuals and institutions and provided that half was to be held in trust for Sylvia's niece, Hetty Robinson. Although Hetty had recently inherited more than one million dollars from her father, she sought her aunt's entire estate under an 1862 will that named her as the sole heir and that provided that no later will should be honored. The executor, Thomas Mandell, claimed that two of the three signatures on the earlier will were traced from an 1864 codicil to the 1863 will, and that even if the earlier will were genuine, the later one applied.³⁰

Both Oliver Wendell Holmes, Sr., Parkman Professor in the Harvard Medical School, and Louis Agassiz, another Harvard professor and one of the world's leading naturalists, examined the contested signatures under a microscope and testified for Robinson that they saw no evidence of tracing.³¹ Mandell countered with testimony from Benjamin Peirce, Professor of Mathematics at Harvard, and his son, Charles Sanders Peirce.³² The Peirces purported to demonstrate that the signatures were forgeries by contrasting the similarities between one of the disputed signatures and its counterpart in the 1864 codicil with the less extensive simi-

²⁸ *Peters*, 534 N.W.2d at 873 (“[A]lthough Wisconsin judges do not evaluate the reliability of scientific evidence, they may restrict the admissibility of such evidence through their limited gatekeeping functions.”).

²⁹ 20 F. Cas. 1027 (C.C.D. Mass. 1868) (No. 11,959). The description of the case that follows is taken largely from Meier & Zabell, *supra* note 17. For a more detailed and comprehensive account of the litigation, see *The Howland Will Case*, *supra* note 17.

³⁰ Meier & Zabell, *supra* note 17, at 497; *The Howland Will Case*, *supra* note 17, at 631–33, 639.

³¹ *The Howland Will Case*, *supra* note 17, at 652–53. According to *Statistical Assessments as Evidence*, *supra* note 19, at 214, Dr. Holmes was “destroyed on cross-examination and forced to admit that he had only first examined the will the previous day.”

³² The Peirces were among the first to contribute to the development of mathematical statistics in the United States, and the younger Peirce later became prominent for his work in philosophy. Meier & Zabell, *supra* note 17, at 497.

larities between the disputed signature and forty-two others on documents written by Sylvia Ann Howland in her later years. C.S. Peirce examined every possible pair of signatures to see how many of the downstrokes in the words "Sylvia Ann Howland" coincided in position and length.³³ He found agreement in approximately one in every five downstrokes. Professor Peirce then testified to an "extraordinary" coincidence in the positions of the thirty downstrokes in the disputed signature and the 1864 signature. He described complete "[c]oincidence of position" as "infallible evidence of design."³⁴ Being a professor of mathematics, Peirce was not content to rest on intuition alone. He insisted that "[t]he mathematical discussion of this subject has never, to my knowledge, been proposed, but it is not difficult; and a numerical expression applicable to this problem, the correctness of which would be instantly recognized by all the mathematicians of the world, can be readily obtained."³⁵ He reasoned that the probability of thirty matches in a given pair of authentic signatures was $(1/5)^{30}$, or "once in 2,666 millions of millions of millions."³⁶ "This number," he added, "far transcends human experience."³⁷ Decided in a century in which scientific and statistical studies received no more scrutiny than any other expert testimony, the admissibility of these calculations was not challenged,³⁸ and even the cross-examination of Peirce was largely ineffectual.³⁹

³³ There are 861 distinct pairs that can be formed from 42 items. Checking 30 downstrokes per item entails 25,830 comparisons. *Id.* at 499.

³⁴ *Id.* (quoting Peirce's testimony).

³⁵ *Id.* (internal quotations omitted).

³⁶ *Id.* This figure is slightly understated, since $5^{30} = 9.31 \times 10^{20}$.

³⁷ He continued: "So vast an improbability is practically an impossibility. Such evanescent shadows of probability cannot belong to actual life. They are unimaginably less than those least things which the law cares not for." *Id.* (quoting Peirce's testimony).

³⁸ From the modern perspective, however, several features of his mathematical analysis are questionable, including the implicit assumption of independence in the matches for the downstrokes. *Id.* at 501. One might expect that agreement in, for example, positions one and three, would make far more likely an agreement in position two. The effect of such dependence would be to increase the probability of thirty matches over that quoted, quite possibly by orders of magnitude. *Id.* Despite such criticisms, Meier and Zabell conclude that "against the background of his era, Peirce's analysis must be judged unusually clear and complete." *Id.* at 503.

³⁹ Peirce's demeanor and reputation as a mathematician must have been intimidating, for there was no inathematical rebuttal. On cross-examination, he

In 1915, however, the New York Court of Appeals held in *People v. Risley*⁴⁰ that even under the relevance-expertise regime, another mathematician's testimony about an alleged forgery was inadmissible in a criminal case. An attorney was charged with fraud in the course of representing a corporate client in a civil matter. Apparently, he had removed a document that had been placed in evidence and typed in the words "the same" to make the meaning more favorable to his client.⁴¹ An expert on typewriters testified that as typed on the document, the six distinct letters in the words "the same" exhibited eleven specific peculiarities.⁴² For example, the "t" was not strictly vertical, but slanted; other letters were missing serifs, and so on. This expert reported that a typewriter removed from Risley's office produced characters with the same peculiarities. A second expert, described by the New York Court of Appeals as "a professor of mathematics in one of the universities of the state," testified that "by the application of the law of mathematical probabilities, the chance of such defects being produced by another typewriting machine was so small as to be practically a negative quantity."⁴³

confessed a lack of special expertise in judging handwriting, but counsel did not cross-examine on the numerical and mathematical parts of the testimony. Nevertheless, the probability computation was not dispositive. The court held that even if the earlier will were Howland's, there was insufficient evidence of the consideration necessary to make binding the instruction to dishonor later wills. The court made no finding on the forgery claim. *Robinson*, 20 F. Cas. at 1032.

⁴⁰ 108 N.E. 200 (N.Y. 1915).

⁴¹ *Id.* at 202.

⁴² *Id.*

⁴³ *Id.* As the court described the testimony:

The witness asserted that, when the facts are ascertained, the application of the law of probabilities to them is a matter of pure mathematics and not of speculation or opinion. He defined the law of probabilities as "a proper fraction expressing the ratio of the number of ways an event may happen, divided by the total number of ways in which it can happen." The various defects claimed to be visible and pointed out by the experts in the specimens of typewriting made upon the defendant's machine . . . were called to the witness' attention by the district attorney, and he was asked to apply the law of mathematical probability thereto. For illustration he was asked: "If it be assumed that it is as probable that any given letter will slant as it is it will not slant, are you able to ascertain what is the probability that the letter 't' in the six letters, 't, h, e, s, a, m,' will slant and the others remain perpendicular?" He answered: "One in 64." Practically the same form of question was put to him in regard to missing serifs on other letters and embracing various features pointed out by the experts, and by a process of compounding these results concerning each of the particular

Over a dissent, the New York Court of Appeals reversed this conviction. The majority questioned the assumption that merely because a letter could slant or not slant, the probability that it would slant is one-half. Observing that the mathematician had no particular knowledge about the frequency of observable defects in typewriters, the court dismissed his statement of the probability because it “was not based upon actual observed data, but was simply speculative.”⁴⁴ In *Robinson*, Peirce had arrived at one-fifth for the probability of two matching strokes by a study of genuine signatures.⁴⁵ In *Risley*, the mathematician had no such empirical foundation for using a value of one-half. Accordingly, the statistical evidence in *Risley* was inadmissible under general principles of relevancy.⁴⁶ As we shall see, even when the doctrinal basis for evaluating scientific testimony became more rigorous, the courts continued to apply the classical relevance-expertise standards to statistical evidence.

B. The Modern Period: Heightened Scrutiny for Scientific Evidence

When a major category of evidence is thought to be unusually prejudicial, ad hoc balancing often crystallizes into more special-

defects the witness was permitted to give his conclusion that the probability of these defects being reproduced by the work of a typewriting machine, other than the machine of defendant, was one in four thousand million, which was, of course, equivalent to a statement that it could never occur.

Id.

⁴⁴ Id. at 203. Of course, the assumption that each peculiarity in the typed characters had a probability of one-half is not the only objectionable feature of the mathematical proof in *Risley*. The “process of compounding” the probabilities pertaining to each letter deserves scrutiny. This is an application of the rule that if two events *A* and *B* are probabilistically independent, then their joint probability is the product of their individual probabilities: $\Pr(AB) = \Pr(A) \times \Pr(B)$. The mathematician in *Risley* arrived at the 1/64 and 1/4,000,000,000 figures by assuming such independence with respect to each characteristic among the six letters of the typewriter keyboard. As a concurring opinion pointed out, however,

the likelihood of similar defects . . . would depend on the dies from which [the type] were made, on the process of manufacture, on the greater likelihood of particular parts, such as serifs, being broken by use, on the material composing the type, on the way in which the machine had been used, and doubtless on many other things.

Id. at 204 (Miller, J., concurring).

⁴⁵ See Meier & Zabell, *supra* note 17, at 499.

⁴⁶ 1 McCormick on Evidence, *supra* note 25, § 210, at 808 & n.12.

ized rules.⁴⁷ For example, evidence of bad character is generally not admissible merely to show a tendency to act wrongly.⁴⁸ Evidence of insurance is not admissible to suggest that the insured might behave carelessly.⁴⁹ In principle, there may be no difference between the pattern of decisions under an ad hoc balancing of probative value and prejudicial effect, but in practice, the presence of a specialized rule reinforces the recognition that the evidence poses special problems. To this extent, it ensures that the evidence receives heightened scrutiny, and it explains the factors that go into this scrutiny. Furthermore, if the rule is not too amorphous, it channels discretion, producing a more uniform and predictable pattern of decisions. If all judges and counsel were perfect and could effortlessly discern the proper outcome of ad hoc balancing, then case-by-case balancing would be ideal. The reality is that unstructured, ad hoc balancing is difficult to do well, and it may be that a cruder but more easily applied rule will produce more consistent outcomes with less effort and little loss in accuracy across all cases.⁵⁰ This is a major argument for categorical rules as opposed to vague standards in many areas of law.⁵¹

Given the pressures for specialized rules of relevance and the perception that scientific evidence poses special problems, it is hardly surprising that courts would come to supplement the relevance-expertise standard with more specific rules that attend to the special features of scientific evidence.⁵² Two forms of additional

⁴⁷ See. *id.* § 185, at 648 ("In certain areas, such as proof of character, comparable situations recur so often that relatively particularized rules channel the exercise of discretion.").

⁴⁸ *Id.* § 186, at 649.

⁴⁹ *Id.* § 201, at 711.

⁵⁰ Cf. Richard A. Epstein, *The Risks of Risk/Utility*, 48 *Ohio St. L.J.* 469 (1987) (defending older common law tort rules over ad hoc balancing of risk and utility); Louis Kaplow, *Rules Versus Standards: An Economic Analysis*, 42 *Duke L.J.* 557 (1992) (offering an economic analysis for whether legal commands should be made as standards or rules); Russell B. Korobkin, *Behavioral Analysis and Legal Form: Rules vs. Standards Revisited*, 79 *Or. L. Rev.* 23 (2000) (discussing the relative merits of rules and standards).

⁵¹ See Kathleen M. Sullivan, *Foreword: The Justices of Rules and Standards*, 106 *Harv. L. Rev.* 22 (1992).

⁵² In Faigman et al., *supra* note 23, Professor David L. Faigman and his colleagues offer another explanation for the introduction of the general acceptance standard, which was the first specialized rule for scientific evidence. Although they recognize that relevance and helpfulness played a role in admitting expert evidence, they locate

scrutiny—general acceptance and scientific soundness—are dominant.

I. General Acceptance: Frye

The general acceptance standard made its debut in the now celebrated case of *Frye v. United States*.⁵³ Alphonse Frye, a young black man in the District of Columbia, was charged with murder. He sought to introduce the testimony of a psychologist, William Moulton Marston, who had administered a systolic blood pressure test to Frye. According to Dr. Marston, the test revealed that Frye was truthful when he denied committing the murder.⁵⁴ Dr. Marston had developed this forerunner of the polygraph test for truthfulness, but it is not clear what he had done to establish its validity.⁵⁵

This testimony could have been excluded under the traditional relevance-expertise standard. Dr. Marston, who was a professor of psychology at Harvard College,⁵⁶ was, of course, qualified to give certain kinds of expert testimony. If his opinion about Frye's veracity was based on a procedure that was not well studied, however, it could have been rejected as too speculative to be of much assis-

the general acceptance standard in the requirement that an expert be qualified. They suggest that the qualifications of the expert were determinative of admissibility: "If the witness was an expert, then his or her opinion testimony was 'entitled' to be admitted as evidence (given, of course, its apparent relevance to the issues to be determined at trial)." *Id.* at 1803. Qualifications, in their view, turned on a "commercial marketplace" test: "If a person could make a living selling his knowledge in the marketplace, then presumably expertise existed." *Id.* at 1804. Because there was no commercial market for forensic science, however, courts were forced to modify the test, and they chose to substitute "[t]he intellectual or professional marketplace" as "a proxy for the commercial marketplace." *Id.* at 1806.

This reconstruction fails to account for expertise that is neither commercial, intellectual, nor scientific. A witness whose expertise was developed as a hobby and for which no commercial market exists would be qualified to give expert testimony, but the admissibility of the proposed testimony would not turn on whether the expert's conclusions were based on methods generally accepted in the field. See *supra* note 23.

⁵³ 293 F. 1013 (D.C. Cir. 1923).

⁵⁴ *Id.*

⁵⁵ But see William M. Marston, *Systolic Blood Pressure Symptoms of Deception*, 2 *J. Experimental Psychol.* 117 (1917) (reporting the results of research on the effects of a consciousness of deception on systolic blood pressure).

⁵⁶ Marston developed the theory of a "specific lie response" while he was a Harvard law student. Years later, he created the comic strip heroine, Wonder Woman, whose golden lasso induced anyone within its circumference to speak the truth. 2 *Who Was Who in America* 347 (1950).

tance to the jury. Indeed, the trial judge, in excluding the testimony, may have been following just this approach.

In affirming the trial court's ruling, the United States Court of Appeals for the District of Columbia observed that "[j]ust when a scientific principle or discovery crosses the line between the experimental and the demonstrable stages is difficult to determine."⁵⁷ This observation is entirely consistent with the traditional approach. A conclusion drawn from a technique that still is "experimental" rather than "demonstrable" may be relevant, but it also may be too insecure to be sufficiently helpful to the jury.

The innovation of *Frye* lies in *how* the court of appeals ascertained whether the technique was too speculative. The court was not content to rely solely on the assertion of the well-qualified expert who had experimented with systolic blood pressure as an indicator of truthfulness; neither was it prepared to inquire directly into whether his work was sufficient to establish the validity of the technique. Rather, it affirmed the exclusion of the evidence on the neoteric ground that *other* psychologists had yet to accept Marston's claim that he could verify honesty by measuring the speaker's blood pressure. Although no previous cases explicitly had held this general acceptance to be indispensable, the court boldly wrote:

Somewhere in this twilight zone [between the experimental and the demonstrable] the evidential force of the principle must be recognized, and while the 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.⁵⁸

The requirement of general acceptance, like any special trustworthiness test, tends to screen out evidence. The *Frye* court offered no reason for imposing this special requirement, but subsequent courts and commentators have filled the gap. As noted above, the rule can be understood as a crystallization of the ad hoc balancing that trial courts are expected to undertake. Ideally, this rule screens out evidence that is superficially impressive but not

⁵⁷ *Frye*, 293 F. at 1014.

⁵⁸ *Id.*

sufficiently probative because it is not scientifically valid. The rule does not ask—or even permit—the court to ascertain scientific validity for itself. Instead, the court must defer to the scientific community, for the rule treats “general acceptance” as a surrogate for validity. By looking to the views of the scientific community, the rule avoids having the judge act like an independent scientist.

Of course, demanding general acceptance as opposed to some lesser degree of support among scientists tends to increase the incidence of “false negatives” (rulings that exclude valid scientific evidence) over “false positives” (rulings that admit invalid scientific evidence). This conservative strategy⁵⁹ has been defended as an appropriate response to the risk that jurors are too credulous of scientific evidence.⁶⁰ Furthermore, waiting until a technique has been generally accepted ensures that it has been widely studied and thus assures that a pool of experts is available to both sides to verify that the technique has been applied properly.

In practice, the objectives of a clear rule—uniformity and predictability—have not been achieved. Courts in different *Frye* jurisdictions have reached contradictory results with respect to the same types of scientific evidence,⁶¹ and it is not obvious that the uniformity achieved under *Frye* is any greater than that which would be obtained with most other plausible rules or standards. Ambiguities as to the propositions that must be generally accepted, the fields in which they must be accepted, the extent to which they must be accepted, and the indicia and proof needed to show their acceptance have made *Frye* disappointingly ductile and frustratingly unpredictable.⁶²

⁵⁹ Fredric I. Lederer, Resolving the *Frye* Dilemma—A Reliability Approach, 26 *Jurimetrics J.* 240, 241 (1986) (“*Frye* tends to be unduly conservative in its effect on the admissibility of novel evidence.”).

⁶⁰ See, e.g., Steven M. Egesdal, Note, The *Frye* Doctrine and Relevancy Approach Controversy: An Empirical Evaluation, 74 *Geo. L.J.* 1769, 1772, 1774 n.26 (1986).

⁶¹ See, e.g., Bert Black et al., Science and the Law in the Wake of *Daubert*: A New Search for Scientific Knowledge, 72 *Tex. L. Rev.* 715, 737–39 (1994) (referring to diverging opinions by courts on voiceprint evidence); Paul C. Giannelli, The Admissibility of Novel Scientific Evidence: *Frye v. United States*, a Half-Century Later, 80 *Colum. L. Rev.* 1197, 1219–21 (1980) (discussing selective application of *Frye* or similar evidentiary demands).

⁶² See, e.g., 1 McCormick on Evidence, supra note 25, § 203, at 727–31; Don E. Walden, Note, *United States v. Downing*: Novel Scientific Evidence and the Rejection of *Frye*, 1986 *Utah L. Rev.* 839, 840–41.

Thus, the use of *Frye* in evaluating statistical assessments has been capricious. Traditionally, *Frye* simply was not perceived as a barrier to statistical testimony.⁶³ Starting in the 1970s, parties in employment discrimination cases brought under Title VII of the Civil Rights Act of 1964 began to make extensive use of statistical expertise.⁶⁴ Early cases involved simple comparisons of proportions,⁶⁵ but as “[t]he [f]loodgates . . . [o]pened,”⁶⁶ more complicated studies were introduced.⁶⁷ Courts discussed standard deviations,⁶⁸ correlation coefficients,⁶⁹ significance levels,⁷⁰ hypothesis tests,⁷¹ Mantel-Haenzel tests,⁷² scattergrams,⁷³ nonlinear regressions,⁷⁴ and

⁶³ See, e.g., *Statistical Assessments as Evidence*, supra note 19, at 220 (citing Symposium on Science and the Rules of Evidence, 99 F.R.D. 188 (William A. Thomas ed., 1983), for the view that “the *Frye* doctrine . . . will almost never limit a statistical expert even if his or her particular statistical theories or methods of analysis are not generally accepted”).

⁶⁴ *Id.* at 102–03.

⁶⁵ E.g., *Washington v. Davis*, 426 U.S. 229, 235–36 (1976) (discussing the disparate impact of a civil service test on African-Americans seeking jobs as police officers); *Griggs v. Duke Power Co.*, 401 U.S. 424, 429–30 (1971) (discussing the disparate impact on African-Americans of a high school diploma requirement and employment tests).

⁶⁶ *Statistical Assessments as Evidence*, supra note 19, at 93.

⁶⁷ *Id.* at 94–102 (describing cases and arguments regarding multiple and logistical regressions). Statistical studies played an important part in Title VII litigation (and in paving the way for the use of statistical expertise in other types of litigation) for a variety of reasons. See *id.* at 102.

⁶⁸ E.g., *EEOC v. Fed. Reserve Bank*, 698 F.2d 633, 651 (4th Cir. 1983) (“If our computation is correct, the standard deviation for pay grade 5 was -1.87 . . .”).

⁶⁹ E.g., *Bernard v. Gulf Oil Corp.*, 890 F.2d 735, 742 (5th Cir. 1989) (“Plaintiffs urge that a correlation coefficient in the .30–.40 range be established as the minimum for proof of a job related test. We decline to establish a bright line cut-off point for the establishment of job-relatedness in testing.”).

⁷⁰ E.g., *Fed. Reserve Bank*, 698 F.2d at 647 (recognizing that the 0.05 level is arbitrary).

⁷¹ E.g., *Moultrie v. Martin*, 690 F.2d 1078, 1082 (4th Cir. 1982) (“[S]tatisticians compare figures through an objective process known as hypothesis testing.”).

⁷² E.g., *Hogan v. Pierce*, 31 Fair Empl. Prac. Cas. (BNA) 115 (D.D.C. 1983).

⁷³ E.g., *Cherry v. Amoco Oil Co.*, 490 F. Supp. 1026, 1028–29, 1030 (D. Ga. 1980) (finding that a sociologist’s scattergram of “credit application acceptance rate” and the proportions of nonwhites residing in zip code regions merely demonstrated that “the computerized grading system [for issuing gasoline credit cards] taken as a whole tends to reject a disproportionate number of persons living in predominantly black areas”).

⁷⁴ E.g., *Craik v. Minn. State Univ. Bd.*, 731 F.2d 465, 476 n.13 (8th Cir. 1984) (noting that logistic regression should have been used, but relying on ordinary least squares regression because neither party explained the difference in the methods).

reverse regressions.⁷⁵ These cases concerned issues such as whether a study that fails to show a disparity that is significant at the 0.05 level could create a prima facie case of disparate impact,⁷⁶ or whether a study that does show a significant difference in salaries but omits certain variables “must be considered unacceptable as evidence of discrimination.”⁷⁷ The opinions and arguments in these cases, however, almost never questioned the admissibility of the evidence. They never suggested that the general acceptance standard or a heightened reliability standard might make the expert’s testimony inadmissible.⁷⁸

Likewise, epidemiological studies in civil cases were admitted with little scrutiny for many years.⁷⁹ In parentage proceedings, courts initially questioned the general acceptance of serological methods,⁸⁰ and they would not admit blood group typing to establish paternity.⁸¹ As the number and power of genetic tests that

⁷⁵ E.g., *Penk v. Or. State Bd. of Higher Educ.*, 36 Empl. Prac. Dec. (CCH) ¶ 35,049 (D. Or. 1985), *aff’d*, 816 F.2d 458 (9th Cir. 1987).

⁷⁶ See *Segar v. Smith*, 738 F.2d 1249, 1282–83 (D.C. Cir. 1984).

⁷⁷ *Bazemore v. Friday*, 751 F.2d 662, 672 (4th Cir. 1984), *rev’d*, 478 U.S. 385 (1986).

⁷⁸ *Bazemore* is a rare case in which the Supreme Court spoke in terms of admissibility of a study said to omit important variables. But the Court did not ask whether the regression conformed to standard statistical practice. Instead, it remarked that a “plaintiff in a Title VII suit need not prove discrimination with scientific certainty,” *id.*, 478 U.S. at 400, and it alluded to the broad principles of relevancy codified in Federal Rules 401 and 403, explaining that “[n]ormally, failure to include variables will affect the analysis’ probativeness, not its admissibility.” *Id.* An accompanying footnote indicates that to be inadmissible for lack of probative value, a regression would have to be grossly inadequate. *Id.* at 400 n.10 (“There may, of course, be some regressions so incomplete as to be inadmissible as irrelevant; but such was clearly not the case here.”).

⁷⁹ See David E. Bernstein, *Frye, Frye, Again: The Past, Present, and Future of the General Acceptance Test*, 41 *Jurimetrics J.* 385, 390 (2001) (noting that until 1988 no court applied *Frye* to a toxic tort case).

⁸⁰ See, e.g., *Huntingdon v. Crowley*, 414 P.2d 382, 390 (Cal. 1966) (holding that lack of general acceptance justified exclusion); *State v. Damm*, 252 N.W. 7, 12 (S.D. 1933) (holding that the lack of medical and scientific agreement justified exclusion).

⁸¹ The test results were admissible to exclude a man as the father, but not to include him as a biologically possible candidate. See *Moore v. McNamara*, 513 A.2d 660, 666 (Conn. 1986) (“The use of blood test results had necessarily been restricted in this way because at one time only a few antigens . . . were known.”); *Flippen v. Meinhold*, 282 N.Y.S. 444, 446 (N.Y. City Ct. 1935) (reporting that “[n]o case has been found in which blood grouping tests have been deemed admissible for the purpose of establishing paternity” and declining to order blood tests for this purpose). This asymmetrical common law rule rested on the disparity in the probative value of the two types of findings. An exclusion was essentially conclusive; but with the limited

could be applied to determine parentage grew, however, the traditional rule began to crumble under the weight of cases⁸² and specialized statutes.⁸³ Laboratories usually accompanied their inclusionary findings with an impressive "probability of paternity"—a statistic that largely went unchallenged. Eventually, some courts restricted the practice,⁸⁴ but the doctrinal basis was not general acceptance. Rather, it was the normal weighing of probative value and prejudicial effect.⁸⁵

Similarly, "[n]ot so long ago, the courts refused to admit either survey or sampling evidence."⁸⁶ Public opinion was not established

number of genetic systems then available, an inclusion was not especially revealing. See *Commonwealth v. English*, 186 A. 298, 300 (Pa. Super. Ct. 1936) (explaining that "in 14 3/4 per cent. of the cases examined the blood grouping test can exonerate, but in no case does it incriminate"); *Flippen*, 282 N.Y.S. at 446. In *Flippen*, the court wrote:

If the test shows a negative result, it would seem to be conclusive proof of nonpaternity, but the positive would simply indicate the possibility of paternity. It would be improper to draw an inference of paternity where merely the possibility is shown; where different inferences may be drawn from a proven fact, no judicial determination may be based thereon.

Id.

⁸² Ira Mark Ellman & David Kaye, *Probabilities and Proof: Can HLA and Blood Group Testing Prove Paternity?*, 54 N.Y.U. L. Rev. 1131, 1131-32 & n.7 (1979).

⁸³ D.H. Kaye & Ronald Kanwisler, *Admissibility of Genetic Testing in Paternity Litigation: A Survey of State Statutes*, 22 Fam. L.Q. 109 (1988).

⁸⁴ See, e.g., *Plemel v. Walter*, 735 P.2d 1209, 1219 (Or. 1987) (imposing three conditions on an expert's testimony regarding the probability of paternity) (discussed in D.H. Kaye, *Plemel as a Primer on Proving Paternity*, 24 Willamette L. Rev. 867 (1988)); D.H. Kaye, *The Probability of an Ultimate Issue: The Strange Cases of Paternity Testing*, 75 Iowa L. Rev. 75, 77-78 (1989) (discussing limitations on the presentation of the "probability of paternity" adopted by various state supreme courts).

⁸⁵ See, e.g., *Plemel*, 735 P.2d at 1218-19; Ellman & Kaye, *supra* note 82, at 1149-50 (discussing the application of Bayes' Theorem). The paternity probability normally is computed according to Bayes' Theorem using a prior probability of one-half. The effort to divorce the paternity probability from the nongenetic evidence in the case poses complications in presenting the genetic evidence fairly. See 1 *Modern Scientific Evidence*, *supra* note 23, §§ 19-1.0 to 2.1, at 748-61; Ellman & Kaye, *supra* note 82, at 1150 (explaining that the method commonly used "is equivalent to supposing that the universe of possible fathers is already reduced to two equally likely suspects before considering the HLA test results").

⁸⁶ Zeisel & Kaye, *supra* note 16, at 101. These authors give the following example:

That attitude led to monstrosities such as *James S. Kirk & Co. v. Federal Trade Commission*, 59 F.2d 179, 181 (7th Cir. 1932),] in which the manufacturer's claim that a soap was based on olive oil was challenged. This earth-shaking issue brought an administrative law judge to Seattle, Washington, where he

through systematic polls but through the testimony of representatives of the public itself—what the law called “public witnesses.”⁸⁷ Thus, in *Irvin v. State*,⁸⁸ the Supreme Court of Florida refused to credit a public opinion survey of community sentiment. Two African-American men were convicted of raping a white woman, but the conviction was set aside after it became clear that the grand jury that returned the indictments had been selected in a discriminatory fashion.⁸⁹ A new grand jury promptly reindicted the men. The NAACP commissioned Elmo Roper, one of the pioneers of American public opinion research, to conduct what was probably the first large-scale survey of public prejudice in a venue. The trial court, however, excluded the research director’s testimony and declined to change the venue.⁹⁰ The trial ended in a verdict of guilt and a sentence of death. The Florida Supreme Court upheld the exclusion of the survey as hearsay and insisted that although a survey might indicate consumer attitudes toward a product, the method was “useless” to “indicate an aroused public against a prospective defendant in a court of justice.”⁹¹ In upholding the refusal to change the venue, the court preferred to rely on “the friendliness of white people for the colored in the community” as

heard, one by one, the testimony of 700 women as to their understanding of the manufacturer’s message.

Id.

⁸⁷ Id. Historically, hearsay has been the major objection to survey evidence. In most large surveys, many persons are employed to do the interviewing or other forms of data collection. Furthermore, when opinion polls are at issue, the individuals whose opinions were sampled are not testifying in court. Various arguments to circumvent or overcome the hearsay rule have been used by courts electing to receive the evidence. See, e.g., *Tex. Aeronautics Comm’n v. Braniff Airways, Inc.*, 454 S.W.2d 199, 203 (Tex. 1970) (holding certain evidence “admissible whether it is considered to be nonhearsay or within the state of mind exception to the hearsay rule”); Hans Zeisel, *The Uniqueness of Survey Evidence*, 45 *Cornell L.Q.* 322, 345–46 (1960) (giving safeguards that would mitigate the dangers of receiving survey evidence).

⁸⁸ 66 So. 2d 288 (Fla. 1953).

⁸⁹ Id. at 289–90 (citing *Shepherd v. State*, 341 U.S. 50, 50 (1951)). Justices Robert Jackson and Felix Frankfurter concurred on the ground that the pretrial publicity had the effect of informing the jurors of an involuntary confession. Id. at 51 (Jackson & Frankfurter, JJ., concurring).

⁹⁰ *Irvin*, 66 So. 2d at 290. Venue had been changed from the county in which the alleged crime was committed to a nearby one. Defendants sought to show that this measure was insufficient. Id.

⁹¹ Id. at 292.

indicated by the testimony of "numerous witnesses" and "the recent construction of an elaborate memorial to a colored soldier."⁹²

In categorically rejecting survey and sampling evidence in *Irvin* and other cases, courts have rarely mentioned *Frye* or any special standards for scientific evidence. Likewise, the later opinions admitting survey results did not maintain that *Frye* was satisfied because social scientists accepted scientific sampling methods to produce opinions. To be sure, modern courts are far more hospitable to survey evidence,⁹³ but the transformation has been traced to other developments.⁹⁴

In criminal cases, the courts have been skeptical of efforts to assign numerical probabilities to events, and often rightly so, but once again, the usual principles of relevance, rather than the special test of general acceptance, have been the vehicle for their expression.⁹⁵ Consider what may be the most famous modern case of statistical testimony introduced to establish a defendant's guilt. In *People v. Collins*,⁹⁶ the Supreme Court of California overturned a conviction because of a contrived (but unchallenged) attempt to show that certain traits of a couple apparently fleeing the scene of a robbery were so uncommon as to be practically conclusive of guilt. Malcolm Collins and his common-law wife Janet had been charged with robbing a woman in an alley in the San Pedro area of Los Angeles. Malcolm was a black man who at one time had worn a beard and mustache and owned a yellow Lincoln; Janet was a caucasian woman with blond hair that she wore in a pony tail. There was no outright confession and no definitive identification of this couple, but a blond woman with her hair in a pony tail was

⁹² *Id.*

⁹³ See generally 1 McCormick on Evidence, *supra* note 25, § 208, at 791-93 (discussing the general admissibility of surveys that are conducted according to certain accepted principles); Susan J. Becker, Public Opinion Polls and Surveys as Evidence: Suggestions for Resolving Confusing and Conflicting Standards Governing Weight and Admissibility, 70 Or. L. Rev. 463 (1991) (discussing the historic treatment of survey evidence and its recent popularity in courts).

⁹⁴ See Zeisel & Kaye, *supra* note 16, at 101 (describing such developments as special statutes admitting Census Bureau reports based on sampling, Judge Wyzanski's *sua sponte* use of sampling in *United States v. United Shoe Machinery Corp.*, 110 F. Supp. 295, 305 (D. Mass. 1953), and an Advisory Committee Note to Federal Rule 703 that (without mentioning *Frye*) speaks approvingly of expert opinions based on sampling).

⁹⁵ 1 McCormick on Evidence, *supra* note 25, § 210, at 808.

⁹⁶ 438 P.2d 33 (Cal. 1968).

seen running from the scene of the robbery and entering a yellow car driven by a bearded and mustached black man.⁹⁷

As in *Risley*, the prosecutor called a college mathematics instructor to the stand and had him assume various values for the frequencies of characteristics like beards, mustaches, interracial couples, and yellow cars. The mathematician then multiplied these assumed values to conclude that the joint probability of all these characteristics in a randomly selected couple would be about 1/12,000,000.⁹⁸

The California Supreme Court reversed the resulting conviction. The opinion, which even sported a mathematical appendix, found at least three errors in the probability testimony: (1) the lack of any evidentiary foundation for the probabilities used by the mathematician; (2) the lack of a foundation for the independence of the events whose probabilities were multiplied together; and (3) the possibility that the jurors were distracted and confused by the mathematical proof.⁹⁹ In *Collins* and other “no-evidence” cases,¹⁰⁰ “the computations have little basis in fact and are presented in the guise of expert analysis.”¹⁰¹ Such calculations are excluded, not be-

⁹⁷ *Id.* at 34.

⁹⁸ *Id.* at 37.

⁹⁹ *Id.* at 38–39. The first was enough to justify reversal by itself. Echoing *Risley*, the court observed that the values used to compute the joint probability were largely speculative:

[T]he prosecution produced no evidence whatsoever showing, or from which it could be in any way inferred, that only one out of every ten cars which might have been at the scene of the robbery was partly yellow, that only one out of every four men who might have been there wore a mustache, that only one out of every ten girls who might have been there wore a ponytail, or that any of the other individual probability factors listed were even roughly accurate.

Id. at 38. Since computations with such numbers have little basis in fact and are dressed in the garb of expert analysis, they should be excluded under the principle that their prejudicial impact outweighs their probative value.

¹⁰⁰ There was no statistical evidence at all in *Collins*. There was testimony about the couple that robbed the woman and testimony about the couple that was arrested, but there was no evidence about the frequency with which the traits of the guilty couple would be found in the population of all couples. Rather, the expert merely told the jury how to evaluate the fact that the Collinses fit the description of the robbers. *Id.* at 40. Thus, as far as statistical evidence goes, *Collins* was actually a “no-evidence” case. There was no evidence of how common or unusual the various incriminating characteristics were.

¹⁰¹ 1 McCormick on Evidence, *supra* note 25, § 210, at 808 (internal citation omitted).

cause the probability model is not generally accepted among statisticians, but “under the principle that their prejudicial impact clearly outweighs their probative value.”¹⁰² Although California was (and remains to this day) a devotee of *Frye*,¹⁰³ the *Collins* opinion contains nary a word about *Frye*, general acceptance, or the way that statisticians usually would estimate the probability of an event like a randomly generated couple sharing all the pertinent traits attributed to the suspects. The opinion is a relevancy opinion, pure and simple.¹⁰⁴

In this regard, *Collins* could not be more different than other opinions of the same court with regard to computations of probabilities of other physical traits attributed to suspects on the basis of biological trace evidence rather than the reports of witnesses. In *People v. Venegas*,¹⁰⁵ a woman was raped in her hotel room. Police sent vaginal swabs and swatches of a bedspread containing two semen stains, along with blood samples from the victim and the defendant, to an FBI laboratory. The FBI reported that defendant’s DNA profile matched the DNA profiles from the swabs and one of the stains, and the FBI added that “the probability of selecting an unrelated individual at random from the Hispanic population with a profile that also matched the samples was approximately 1 in 31,000.”¹⁰⁶ After a hearing on the general acceptance of the procedure for arriving at this figure, the trial court admitted testimony that “the probability of another [randomly selected] person having the DNA profile found in defendant’s blood sample was 1 in 65,000.”¹⁰⁷ Both the state court

¹⁰² *Id.*

¹⁰³ *People v. Leahy*, 882 P.2d 321, 323, 331 (Cal. 1994).

¹⁰⁴ Well, not so simple. The court’s appendix is a masterpiece of obfuscation, serving to make a simple point mysterious, and thus buttressing the court’s dubious claim that it was unreasonable to expect defense counsel to recognize and demonstrate the glaring infirmities in the computations provided in *Collins*. See *Collins*, 438 P.2d at 42–43.

¹⁰⁵ 954 P.2d 525 (Cal. 1998).

¹⁰⁶ *Id.* at 530.

¹⁰⁷ *Id.* at 543. The smaller frequency of 1 in 65,000 was calculated with “the modified ceiling method” for combining the frequencies of each DNA allele as estimated with “floating bins.” *Id.* In its original report on the case, the FBI used a “fixed bin” approach. *Id.* The fixed bin approach also led to estimated frequencies of one in 53,000 for the caucasian population and one in 225,000 for the black population. *Id.* For explanations of the binning procedures, see, for example, NRC Report, *supra*

of appeals and supreme court agreed that the method for arriving at the probability had to be generally accepted in the scientific community. Ultimately, the California Supreme Court held that the number came from a computational procedure that was not generally accepted because of an inconsistency between the statistical criterion used in declaring a match and the one used in estimating the frequencies of matching alleles.¹⁰⁸ Likewise, in *People v. Soto*,¹⁰⁹ the California Supreme Court looked to general acceptance in “the relevant scientific community of population geneticists” to conclude that “statistical calculations” for DNA types using “the unmodified product rule” met the *Frye* standard for admissibility.¹¹⁰

One explanation for the unexplained shift from relevancy in *Collins* to general acceptance in *Soto* and *Venegas* might be that the probability computations in the DNA cases could not be dismissed as utterly devoid of an empirical foundation or a theory that might justify the independence assumption. Forensic scientists had compiled some data as to the frequencies of the various alleles that comprise the more complex genotypes, and geneticists had some experience and an ample theoretical framework to draw on in inferring genotype frequencies. Although some defendants vainly argued that *Collins* precluded any multiplication of probabilities,¹¹¹ the DNA computations simply could not be dismissed as manifestly erroneous and hence irrelevant.¹¹² Consequently, a further argument, such as the lack of general acceptance of the probability calculations, was necessary if defendants were to block the evidence. Nevertheless, DNA cases stand out as the only instance in

note 16, at 142–45; David H. Kaye, DNA Evidence: Probability, Population Genetics, and the Courts, 7 Harv. J.L. & Tech. 101, 122–51 (1993).

¹⁰⁸ For an exposition of the FBI's error, see 1 *Modern Scientific Evidence*, supra note 23, § 15-4.2, at 234–36 (Supp. 2000).

¹⁰⁹ 981 P.2d 958 (Cal. 1999).

¹¹⁰ *Id.* at 960.

¹¹¹ See, e.g., Motion to Exclude DNA Evidence at *52–53, *People v. Simpson*, 1994 WL 568647 (Cal. Super. Ct. Oct. 5, 1994) (No. BA097211); David H. Kaye, The DNA Chronicles: Is Simpson Really Collins?, O.J. Simpson Case Commentaries, Nov. 1, 1994, available at 1994 WL 592117.

¹¹² See, e.g., Kaye, supra note 107, at 127–51 (arguing that concern over population structure is misplaced when the reference population is a broad racial group).

which courts in *Frye* jurisdictions have responded to criminal "probability evidence" with a *Frye* analysis.¹¹³

2. *Relevancy-Plus: The Road to Daubert*

The general acceptance standard never was popular with evidence scholars,¹¹⁴ and by the 1970s and 1980s, more and more courts abandoned it in favor of various substitutes.¹¹⁵ For example, in *United States v. Williams*,¹¹⁶ the government recorded telephone conversations initiated by an undercover police officer offering to buy heroin. At trial, it introduced a spectrographic analysis to prove that the voice on the recording was the defendant's. In upholding the admission of this testimony, the Court of Appeals for the Second Circuit refused to apply general acceptance as a "universal[] litmus test for the general admissibility of all 'scientific' evidence."¹¹⁷ Instead, the court recited the usual features of relevancy¹¹⁸ and concentrated on "reliability."¹¹⁹ The Second Circuit concluded that the technique possessed the requisite reliability to warrant admission in light of the extent of its acceptance in the scientific community and "the potential rate of error."¹²⁰

Some years later, in *United States v. Downing*,¹²¹ the Court of Appeals for the Third Circuit expounded at length on this notion

¹¹³ Cases from other *Frye* jurisdictions that do not inquire into general acceptance of probability calculations for trace evidence include *State v. Garrison*, 585 P.2d 563, 566 (Ariz. 1978) (allowing a dubious probability estimate for a bite mark), and *State v. Kim*, 398 N.W.2d 544, 547, 549 (Minn. 1987) (allowing ABO and PGM typing, but excluding testimony as to the probability of exclusion as unfairly prejudicial).

¹¹⁴ See, e.g., 1 McCormick on Evidence, supra note 25, § 203, at 727-31.

¹¹⁵ Id. § 203, at 730 & n.35; Giannelli, supra note 61, at 1228-33.

¹¹⁶ 583 F.2d 1194 (2d Cir. 1978).

¹¹⁷ Id. at 1197.

¹¹⁸ Id. at 1198 ("[T]he established considerations applicable to the admissibility of evidence come into play, and the probativeness, materiality, and reliability of the evidence, on the one side, and any tendency to mislead, prejudice, or confuse the jury on the other, must be the focal points of inquiry.").

¹¹⁹ Id. at 1198-99. A better term would have been "validity." See, e.g., 1 McCormick on Evidence, supra note 25, § 203.

¹²⁰ *Williams*, 583 F.2d at 1198-99 (listing several reasons to think that the error rate should be small). In retrospect, it seems clear that the court was far too sanguine in its assessment of "reliability." See, e.g., 1 McCormick on Evidence, supra note 25, § 207, at 789-90 (discussing the questionable validity and reliability of spectrogram comparisons).

¹²¹ 753 F.2d 1224 (3d Cir. 1985).

that the admissibility of scientific evidence requires “a quantum of reliability beyond that required to meet a standard of bare logical relevance,” and explained that this condition can be fulfilled even when “the principles underlying the evidence have not become ‘generally accepted’ in the field to which they belong.”¹²² The defendant, who was convicted for fraud on the basis of eyewitness identifications, was precluded from presenting a psychologist to testify to experiments on the sources of eyewitness error. The court of appeals remanded to permit the district court to reconsider its ruling in light of the criteria for ascertaining admissibility articulated in this repudiation of *Frye*. Under *Downing*, “reliability” is “a critical element of admissibility,”¹²³ and the “reliability inquiry”¹²⁴ can probe the “degree of acceptance within [the scientific] community,”¹²⁵ the “existence of a specialized literature dealing with the technique,”¹²⁶ and the “rate of error.”¹²⁷ In addition, *Downing* called on the district court to inquire into “another aspect of relevancy”—“fit,” that is, “whether expert testimony proffered in the case is sufficiently tied to the facts of the case that it will aid the jury in resolving a factual dispute.”¹²⁸

As *Williams* and *Downing* indicate,¹²⁹ the major emergent alternative to *Frye* looked to the relevance of the proposed scientific testimony but demanded something more—relevance plus a certain extra trustworthiness, accuracy, or fit beyond that needed to admit nonscientific testimony.¹³⁰ Statistical evidence, however, was

¹²² Id. at 1235.

¹²³ Id. at 1238.

¹²⁴ Id. (“The reliability inquiry that we envision is flexible and may turn on a number of considerations . . .”).

¹²⁵ Id.

¹²⁶ Id.

¹²⁷ Id. at 1239.

¹²⁸ Id. at 1242.

¹²⁹ The approach exemplified in these cases was presaged and supported by many commentators. See, e.g., 3 Jack B. Weinstein et al., *Weinstein’s Evidence* ¶ 702[03], at 702-43 to 702-44 (1996); Mark McCormick, *Scientific Evidence: Defining a New Approach to Admissibility*, 67 *Iowa L. Rev.* 879 (1982).

¹³⁰ Arguably, *State v. Brown*, 687 P.2d 751 (Or. 1984), is an exception. There, the Oregon Supreme Court dismissed both “general acceptance” and “reasonably reliable” as the standards for admitting scientific evidence—in that instance, polygraph testing. Id. at 759. It also “found the [ordinary] relevancy test not satisfactory because it is too nebulous.” Id. Instead, the court relied on “the ‘relevancy’ test [as] strengthened by consideration of . . . seven factors set forth . . . in

rarely held to this standard. The "relevancy-plus" jurisdictions, like the *Frye* jurisdictions, either admitted statistical studies with little comment or excluded them as too flawed to satisfy the more general balancing standard of Federal Rule of Evidence 403.¹³¹ With the Supreme Court's opinion in *Daubert*, however, this situation would change. The courts would not necessarily demand more of statistics, but the doctrinal machinery for processing scientific evidence no longer would remain idle or overlooked when statistical studies were offered.

3. *Scientific Soundness: Daubert*

After many years of refusing to examine the issue of the admissibility of scientific evidence,¹³² the Supreme Court granted certiorari in *Daubert* to consider whether the general acceptance standard survived the enactment of the Federal Rules of Evidence. In *Daubert*, two young children born with deformed limbs and their parents sought damages against the manufacturer of Bendectin, a prescription drug taken by the boys' mothers to treat nausea and vomiting during pregnancy. The plaintiffs' case foundered when they were unable to point to any published epidemiological studies concluding that Bendectin causes limb reduction defects. The district court granted the defendant's motion for summary judgment on the ground that the plaintiffs had failed to establish a genuine issue of material fact regarding causation. As summarized by the Court of Appeals for the Ninth Circuit:

Plaintiffs' evidence of causation consisted primarily of expert opinion based on in vitro and in vivo animal tests, chemical structure analyses and the reanalysis of epidemiological studies. Among the contrary evidence proffered by Merrell Dow was

3 Weinstein's Evidence 702[03], pp. 702-15 to 702-21 (1982)" to "provide structure and guidance." *Brown*, 687 P.2d at 759. Consideration of these factors does not necessarily imply that scientific evidence must be especially probative; nonetheless, the factors lend themselves to being applied so as to demand greater reliability than would be needed to admit nonscientific evidence. Thus, *Brown* and cases like it probably are best classified as part of the "relevancy-plus" camp.

¹³¹ Fed. R. Evid. 403 (stating that "[a]lthough relevant, evidence may be excluded if its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury, or by considerations of undue delay, waste of time, or needless presentation of cumulative evidence").

¹³² E.g., *Christophersen v. Allied-Signal Corp.*, 503 U.S. 912 (1992) (denying certiorari for this case involving the standard of admissibility for an expert opinion).

the affidavit of a physician and epidemiologist who reviewed all of the available literature on the subject, which included more than 30 published studies involving over 130,000 patients, and concluded that no published epidemiological study had demonstrated a statistically significant association between Bendectin and birth defects. Plaintiffs do not challenge this summary of the published record.¹³³

The trial court in *Daubert* excluded all four categories of the plaintiffs' evidence—so-called “structure-activity studies,”¹³⁴ in vitro or animal cell experiments,¹³⁵ in vivo or live animal research,¹³⁶ and reanalysis of the epidemiological data.¹³⁷ These rulings on admissibility were based on two lines of reasoning. First, the district and circuit courts held that, absent scientific understanding of the cause of the birth defects in question, causation may only be shown through epidemiological evidence.¹³⁸ Second, both courts refused to allow the recalculated epidemiological data offered by plaintiffs' experts because, unlike the studies “rejected by [the plaintiffs' experts, which] had been published in peer-reviewed scientific

¹³³ *Daubert v. Merrell Dow Pharms., Inc.*, 951 F.2d 1128, 1129 (9th Cir. 1991) (emphasis omitted), vacated by 509 U.S. 579 (1993).

¹³⁴ Five of plaintiffs' experts were willing to opine that the similarity of chemical structure between ingredients in Bendectin and known teratogens constituted evidence that Bendectin caused birth defects. *Daubert v. Merrell Dow Pharms., Inc.*, 727 F. Supp. 570, 574–75 (S.D. Cal. 1989), aff'd, 951 F.2d 1128 (9th Cir. 1991), vacated by 509 U.S. 579 (1993); Brief for Petitioners at 4–5, *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579 (1993) (No. 92-102) [hereinafter Petitioners' Brief].

¹³⁵ These experiments showed that the ingredients in Bendectin caused minor DNA damage to cells in culture or inhibited limb bud cell differentiation. *Turpin v. Merrell Dow Pharms., Inc.*, 959 F.2d 1349, 1358 (6th Cir. 1992); *Brock v. Merrell Dow Pharms., Inc.*, 874 F.2d 307, 313–14 (5th Cir. 1989), modified by 884 F.2d 166; Petitioners' Brief, supra note 134, at 4.

¹³⁶ These were studies of fetal defects in pregnant animals such as rats, rabbits, and monkeys. See *Turpin*, 959 F.2d at 1358–59; *Brock*, 874 F.2d at 314; Petitioners' Brief, supra note 134, at 4.

¹³⁷ After some of the reported instances of injury or noninjury following use of the drug were reclassified, the correlation between drug-usage and birth defects became statistically significant. *Daubert*, 951 F.2d at 1130–31 (discussing plaintiffs' “reanalysis of epidemiological studies” in general terms); *Richardson v. Richardson-Merrell, Inc.*, 857 F.2d 823, 831 (D.C. Cir. 1989) (noting that one expert obtained “what he deems a statistically significant result” only “by recalculating”).

¹³⁸ *Daubert*, 727 F. Supp. at 572, 575; *Daubert*, 951 F.2d at 1131.

journals," the plaintiffs' experts had "neither published [their] recalculations nor offered them for [peer] review."¹³⁹

The Supreme Court unanimously held that the lower courts had applied the wrong standard for the admissibility of scientific evidence. In an opinion by Justice Harry A. Blackmun, the Court proclaimed that the "austere [general acceptance] standard, absent from and incompatible with, the Federal Rules of Evidence, should not be applied in federal trials."¹⁴⁰ In reaching this conclusion, the Court made no effort to analyze the substance or merits of the general acceptance standard, but relied instead on the fact that neither the wording nor the drafting history of the rules of evidence evinced "any clear indication that Rule 702 or the Rules as a whole were intended to incorporate a 'general acceptance' standard."¹⁴¹

Having jettisoned general acceptance as "the exclusive test for admitting expert scientific testimony,"¹⁴² the Court adopted the richer and more flexible¹⁴³ "relevancy-plus" standard already employed in many jurisdictions.¹⁴⁴ It announced that as the gatekeeper

¹³⁹ *Daubert*, 727 F.Supp. at 573, 575-76 (emphasis omitted) (internal quotation omitted) (quoting *Richardson*, 857 F.2d at 831). Most other circuits had reached the same result on the admissibility of testimony that Bendectin is a teratogen. These opinions, however, are better understood as addressing the sufficiency of the scientific evidence on causation to support a verdict for plaintiffs. See Michael D. Green, Expert Witnesses and Sufficiency of Evidence in Toxic Substances Litigation: The Legacy of *Agent Orange* and Bendectin Litigation, 86 Nw. U. L. Rev. 643 (1992); Samuel R. Gross, Substance and Form in Scientific Evidence: What *Daubert* Didn't Do, in *Reforming the Civil Justice System* 234 (Larry Kraemer ed., 1996); cf. Joseph Sanders, The Bendectin Litigation: A Case Study in the Life Cycle of Mass Torts, 43 Hastings L.J. 301 (1992) (describing how law and science have interacted and advanced through the course of the Bendectin litigation).

¹⁴⁰ *Daubert*, 509 U.S. at 589.

¹⁴¹ *Id.* at 588. This aspect of the majority opinion is consistent with the Court's often mechanical approach to interpreting the Federal Rules of Evidence. See generally Glen Weissenberger, The Supreme Court and the Interpretation of the Federal Rules of Evidence, 53 Ohio St. L.J. 1307 (1992) (criticizing this jurisprudence). The more convincing view is that the rules left the viability of the general acceptance standard open to further common law development. See, e.g., *Downing*, 753 F.2d at 1235 (concluding that "the Federal Rules of Evidence neither incorporate it nor repudiate it"); *Christophersen v. Allied-Signal Corp.*, 939 F.2d 1106, 1111, 1115-16 (5th Cir. 1991) (en banc); Paul Giannelli, *Daubert: Interpreting the Federal Rules of Evidence*, 15 Cardozo L. Rev. 1999, 2002, 2016-19 (1994).

¹⁴² *Daubert*, 509 U.S. at 589.

¹⁴³ *Id.* at 594 (citation omitted) ("The inquiry envisioned by Rule 702 is . . . a flexible one.").

¹⁴⁴ See *supra* Section I.B.2.

of evidence, “the trial judge must ensure that any and all scientific testimony or evidence admitted is not only relevant, but reliable.”¹⁴⁵ This “evidentiary reliability,” as the Court put it, presumes “scientific knowledge”¹⁴⁶—the proffered testimony must be “ground[ed] in the methods and procedures of science.”¹⁴⁷ In a further elaboration, the Court suggested that this “reliability” determination “entails a preliminary assessment of whether the reasoning or methodology underlying the testimony is scientifically valid and . . . properly can be applied to the facts in issue.”¹⁴⁸ This, in turn, depends on such things as “whether it can be (and has been) tested,” “whether the theory or technique has been subjected to peer review and publication,” “the known or potential rate of error,” and the “degree of acceptance within [a relevant scientific] community.”¹⁴⁹

Moreover, the Court suggested, a showing of scientific validity is not enough, for “Rule 702’s ‘helpfulness’ standard requires a valid scientific connection to the pertinent inquiry as a precondition to admissibility.”¹⁵⁰ Drawing directly on *Downing*, the Court observed that “whether expert testimony proffered in the case is sufficiently tied to the facts of the case . . . has been aptly described by Judge Becker as one of ‘fit.’”¹⁵¹ As a logical matter, however, the fit requirement is superfluous. “Purpose” is already built into the definition of “validity.” For example, the LSAT has been shown to

¹⁴⁵ *Daubert*, 509 U.S. at 589. Chief Justice William H. Rehnquist and Justice John Paul Stevens dissented from this part of the opinion. They would have decided only the *Frye* issue and left “the further development of this important area of the law to future cases.” *Id.* at 601 (Rehnquist, C.J., concurring in part and dissenting in part). The opinion of the Chief Justice, joined by Justice Stevens, stated: “I do not doubt that Rule 702 confides to the judge some gatekeeping responsibility in deciding questions of the admissibility of proffered expert testimony.” *Id.* at 600 (Rehnquist, C.J., concurring in part and dissenting in part). With some justification, however, they complained that the majority’s pronouncements were “general, . . . vague and abstract.” *Id.* at 598 (Rehnquist, C.J., concurring in part and dissenting in part).

¹⁴⁶ *Id.* at 590–91.

¹⁴⁷ *Id.* at 590.

¹⁴⁸ *Id.* at 592–93. Likewise, the Court cautioned that “to qualify as ‘scientific knowledge,’ an inference or assertion must be derived by the scientific method. Proposed testimony must be supported by appropriate validation—i.e., ‘good grounds,’ based on what is known.” *Id.* at 590.

¹⁴⁹ *Id.* at 593–94.

¹⁵⁰ *Id.* at 591–92.

¹⁵¹ *Id.* at 591.

be valid for the purpose of predicting grades in the first year of law school.¹⁵² The test is not valid for predicting monetary success as a lawyer.¹⁵³ But even if “fit” is implicit in scientific validity, the discussion in *Daubert* is an important reminder that “scientific validity for one purpose is not necessarily scientific validity for other, unrelated purposes.”¹⁵⁴

The impact of *Daubert* far exceeds its substance. The opinion adds little to the relevancy-plus standard developed in the decades preceding it.¹⁵⁵ Nevertheless, lower courts were stunned. One district court exclaimed that “[t]he rules governing the admissibility of expert testimony have recently undergone dramatic change.”¹⁵⁶ On the remand in *Daubert* itself, Judge Alex Kozinski spoke of the “Brave New World”¹⁵⁷ that the court faced.¹⁵⁸ Invoking the metaphor of “gatekeeping”—hardly a new concept in the law of

¹⁵² Linda F. Wightman, Predictive Validity of the LSAT: A National Summary of the 1990–1992 Correlation Studies 9–10 (1993); cf. Linda F. Wightman & David G. Muller, An Analysis of Differential Validity and Differential Prediction for Black, Mexican-American, Hispanic, and White Law School Students 11–13 (1990); David Kaye, Searching for Truth About Testing, 90 Yale L.J. 431 (1980) (book review).

¹⁵³ One always can attack the evidence strictly in terms of the proof of scientific validity (as indicated by reference to the specific *Daubert* factors) for a particular purpose. The Court’s example of a valid scientific knowledge that should be excluded for lack of fit illustrates this point. Justice Blackmun wrote that:

The study of the phases of the moon, for example, may provide valid scientific “knowledge” about whether a certain night was dark, and if darkness is a fact in issue, the knowledge will assist the trier of fact. However (absent creditable grounds supporting such a link), evidence that the moon was full on a certain night will not assist the trier of fact in determining whether an individual was unusually likely to have behaved irrationally on that night.

Daubert, 509 U.S. at 591. One can say, as the Court does, that knowledge of darkness does not “fit” the facts of the case (which involve irrationality). Or, one can proceed exclusively under the “reliability” prong of *Daubert* by saying that the theory that the full moon causes irrational behavior lacks validity.

¹⁵⁴ *Id.*

¹⁵⁵ Indeed, the similarities in the standard for admissibility described in *Daubert* and the standard articulated by Judge Becker in *Downing* are striking.

¹⁵⁶ *Ohio v. Louis Trauth Dairy, Inc.*, 925 F. Supp. 1247, 1250 (S.D. Ohio 1996).

¹⁵⁷ *Daubert v. Merrell Dow Pharms., Inc.*, 43 F.3d 1311, 1315 (9th Cir. 1995). The task of ascertaining relevance and reliability, he suggested, was “complex and daunting,” “difficult,” “uncomfortable,” and “heady.” *Id.* at 1315–16.

¹⁵⁸ In the end, the Ninth Circuit adhered to its initial decision on the ground that even if the reanalyses were admissible under the principles articulated by the Supreme Court, they showed too small an association between exposure and birth defects to establish specific causation. *Id.* at 1320–22.

evidence¹⁵⁹—courts began to re-examine seemingly settled results as to the admissibility of many forms of scientific testimony.¹⁶⁰ Some scientific evidence was admitted more readily,¹⁶¹ but much was reviewed with a newfound skepticism and a sense of disquiet. In particular, pretrial motions to exclude statistical testimony became commonplace.¹⁶² Along with the shift in focus from weight to

¹⁵⁹ See, e.g., Margaret A. Berger, *Expert Testimony: The Supreme Court's Rules, Issues in Sci. & Tech.*, Summer 2000, at 57, 58. As that article explains:

Perhaps the most significant part of *Daubert* is the Court's anointment of the trial judge as the "gatekeeper" who must screen proffered expertise to determine whether the relevancy and reliability prongs are met. Although there was nothing particularly novel about a trial judge having the power to exclude inappropriate expert testimony, *Daubert* stressed that the trial court has an obligation to act as gatekeeper even though some courts would rather have left this task to the jury, especially when the screening entailed complex scientific issues.

Id.

¹⁶⁰ E.g., *United States v. Cordoba*, 104 F.3d 225, 227 (9th Cir. 1997) (concluding that "our per se rule excluding the admission of unstipulated polygraph evidence was effectively overruled by *Daubert*"); *United States v. Posado*, 57 F.3d 428, 429 (5th Cir. 1995) (concluding that "the rationale underlying this circuit's per se rule against admitting polygraph evidence did not survive *Daubert*").

¹⁶¹ See, e.g., *United States v. Bonds*, 12 F.3d 540, 562 (6th Cir. 1993) (upholding the admission of DNA evidence notwithstanding seemingly substantial scientific controversies).

¹⁶² See, e.g., *Munoz v. Orr*, 200 F.3d 291, 301 (5th Cir.) (plaintiffs' expert's statistical analysis properly excluded as unreliable under *Daubert* for problems ranging "from particular miscalculations to his general approach to the analysis" including tables that did not add to anywhere near 100% and failure to perform a regression analysis and thereby account for pertinent variables), cert. denied, 531 U.S. 812 (2000); *Johnson Elec. N.A., Inc. v. Mabuchi Motor Am. Corp.*, 103 F. Supp. 2d 268, 282-86 (S.D.N.Y. 2000) (invoking *Daubert*, *Joiner*, and *Kumho* to exclude a "speculative" and "preposterous" econometric model for estimating demand in a patent infringing case "despite its dazzling sheen of erudition and meticulous methodology"); *In re Polypropylene Carpet Antitrust Litig.*, 2000-2 Trade Cas. (CCH) ¶ 72,981, at 88,341, 88,349 (N.D. Ga. Apr. 27, 2000) (denying motion to permit interlocutory review of pretrial ruling to admit economist's testimony about prices based on regression said to omit an important variable); *Allapattah Servs., Inc. v. Exxon Corp.*, 61 F. Supp. 2d 1335, 1353 (S.D. Fla. 1999) (admitting econometric testimony under modified *Daubert* analysis); *In re Indus. Silicon Antitrust Litig.*, No. 95-2104, 1998 WL 1031507, at *2-3 (W.D. Pa. Oct. 13, 1998) (holding that a "before and after" regression analysis satisfies *Daubert* and *Bazemore*); *Estate of Bud Hill v. Conagra Poultry Co.*, No. CIV.A.4:94CV0198-HLM, 1997 WL 538887, at *6-9 (N.D. Ga. Aug. 25, 1997) (denying motion to exclude economist's regression study to determine whether chickens were misweighed); *Diehl v. Xerox Corp.*, 933 F. Supp. 1157, 1167-68 (W.D.N.Y. 1996) (denying motion to exclude simple comparisons rather than regressions to show disparate impact); *Newport Ltd. v. Sears, Roebuck &*

admissibility came a series of problems involving the structure, reach, and appellate review of the heightened scrutiny of scientific, expert testimony—and two more Supreme Court opinions on these issues.

C. The Puzzles of Strict Scrutiny

1. The Boundary Problem

If scientific evidence must clear a hurdle that does not block the path of other expert testimony, the problem of demarcating boundaries arises. What evidence counts as “scientific” for the purpose of *Frye*, *Daubert*, or any other such standard? Advocates have implored courts to apply heightened scrutiny to a myriad of claims. Some items, such as agglutination or electrophoresis of blood or the spectrographic analysis of voices, seem indisputably “scientific.” Courts have not hesitated to apply the special standards to testimony about such technologies.¹⁶³ Other testimony, such as the opinion of a psychiatrist that a person’s will is overborne by a compulsion to gamble,¹⁶⁴ seems less easy to classify. In these borderline cases, courts have reached apparently conflicting results; few opinions have provided clear or comprehensive explanations of how the line was drawn.¹⁶⁵

Co., No. 86–2319, 1995 WL 328158, at *2 (E.D. La. May 30, 1995) (denying *Daubert* motion to exclude calculations of lost profits involving “the absorption rate of industrial park land” based on multiple regression model because regression in general and as used to estimate lost profits in other contexts is accepted and valid). Of course, some of the testimony excluded as a result of pretrial “*Daubert* motions” might have been excluded at trial in the pre-*Daubert* era, but the growth in pretrial attacks indicates the enhanced sensitivity of lawyers and courts to the issue of admissibility.

¹⁶³ E.g., Michael R. Flaherty, Annotation, Admissibility, in Criminal Cases, of Evidence of Electrophoresis of Dried Evidentiary Bloodstains, 66 A.L.R.4th 588, 605–09 (1988); cf. *People v. Coleman*, 759 P.2d 1260, 1277 (Cal. 1988) (holding that a positive hemostick test for the presence of blood was improperly admitted when the prosecution did not establish that the hemostick was a generally accepted method for detecting blood).

¹⁶⁴ *United States v. Lewellyn*, 723 F.2d 615, 619–20 (8th Cir. 1983) (holding that a defendant claiming insanity due to pathological gambling must show that the mental health community generally accepts the principles underpinning his theory).

¹⁶⁵ David E. Bernstein, *The Science of Forensic Psychiatry and Psychology*, 2 *Psychiatry, Psychol. & L.* 75, 78 (1995).

Statistical evidence, it seems, is such a borderline case. For instance, in *Ohio v. Louis Trauth Dairy, Inc.*,¹⁶⁶ a federal district court noted that econometrics and statistics were simply methods applied to produce knowledge in substantive disciplines. As such, the court concluded that “[n]either economics or statistics seems to completely qualify as ‘scientific knowledge’” for purposes of *Daubert*.¹⁶⁷ In the textualist style of *Daubert*, this opinion seeks to resolve the boundary problem by asking what scientists (rather than statisticians) “know.” Statistical reasoning, however, is crucial to most scientific inquiry—indeed, some would say that it is the essence of all inductive scientific reasoning. It is required of (although not always mastered by) students of the “hard” as well as the softer sciences. Although statistical modeling is as much art as science,¹⁶⁸ statistical techniques and tests have well-defined mathematical properties described in an active research literature. In a word, it is not a misnomer to speak of “statistical science.”¹⁶⁹ From this perspective, it would seem that the focus in *Trauth Dairy* on whether statistical expertise is a substantive, empirical science like physics, astronomy, or psychology misses the mark.

Yet, this conclusion may be too facile. What, one might well ask, are the unstated criteria being used to separate “science” from other knowledge? At first glance, philosophical studies of the nature and structure of scientific theories might seem to hold the key

¹⁶⁶ 925 F. Supp. 1247 (S.D. Ohio 1996).

¹⁶⁷ *Id.* at 1252. Nevertheless, the court followed the rule in the Sixth Circuit, which anticipated the result in *Kumho*, to conclude that various statistical studies said to indicate a price-fixing conspiracy were admissible under *Daubert* as “modified in the case of social science or other non-scientific expertise.” *Id.* The federal District Court for the Northern District of Alabama excluded similar studies by the same expert in an unrelated case, but in *City of Tuscaloosa v. Harcros Chemicals, Inc.*, 158 F.3d 548 (11th Cir. 1998), the Eleventh Circuit held most of that evidence admissible under *Daubert*. *Id.* at 563.

¹⁶⁸ See, e.g., Lincoln E. Moses, *The Reasoning of Statistical Inference*, in *Perspectives on Contemporary Statistics* 107, 117–18 (David C. Hoaglin & David S. Moore eds., 1992).

¹⁶⁹ The phrase is the title of a respected journal. See also *Statistical Science in the Courtroom*, *supra* note 19 (discussing various uses of statistics in litigation). Of course, admissibility of particular applications of the “science” of statistics depends on much more than the existence of a procedure with well-defined mathematical properties. See *infra* Part II.

to this puzzle. Indeed, the *Daubert* Court started down this road¹⁷⁰ when it cited Sir Karl Popper's criteria for distinguishing science from metaphysics.¹⁷¹ Nevertheless, the basis for drawing a line between expert scientific evidence and other expert testimony is not to be found in abstract definitions of "science." The writings of David Hume, Immanuel Kant, A.J. Ayer, Sir Karl Popper, Thomas Kuhn, and many other philosophers or historians provide brilliant insights into the nature of scientific knowledge, but they do not speak directly to the legal issues.¹⁷² Enriching as the philosophical literature on the nature and aims of science might be, it is unlikely to be of great assistance in deciding when a special test for scientific evidence should be applied. The reason, as Justice Oliver Wendell Holmes once remarked, is that "[a] word is . . . the skin of a living thought."¹⁷³ Words are the visible surface of rules that are designed to achieve certain goals. Abstract definitions may or may not fit these goals.¹⁷⁴

Thus, a functional inquiry, rather than a review of the philosophical literature, the encyclopedia, or the dictionary is required. The rules of evidence, whether derived from the common law or a code, are designed to perform certain functions, and the *raison d'être* of a special hurdle for scientific evidence is that this particular evidence poses special problems. When these problems are not present, heightened scrutiny is not justified and may well be coun-

¹⁷⁰ See generally Scott Brewer, *Scientific Expert Testimony and Intellectual Due Process*, 107 *Yale L.J.* 1535, 1547-50 (1998) (describing the *Daubert* Court's attempts to discern the philosophical boundaries of Federal Rule 702's reference to "scientific knowledge").

¹⁷¹ *Daubert*, 509 U.S. at 593.

¹⁷² But see Kenneth R. Foster & Peter W. Huber, *Judging Science: Scientific Knowledge and the Federal Courts* (1997) (collecting literature on how science is practiced to identify criteria for ascertaining knowledge that is scientific and that can be relied upon).

¹⁷³ *Towne v. Eisner*, 245 U.S. 418, 425 (1918).

¹⁷⁴ Some writers, harking back to an older tradition, define "science" as "classification" or "a coherent, systematic body of knowledge, combining particular facts with general principles." Mike Townsend, *Implications of Foundational Crises in Mathematics: A Case Study in Interdisciplinary Legal Research*, 71 *Wash. L. Rev.* 51, 58 (1996); Harold J. Berman & Charles J. Reid, Jr., *The Transformation of English Legal Science: From Hale to Blackstone*, 45 *Emory L.J.* 437, 440 (1996). Perhaps such a broad conception of science, which could encompass religion, law, and art history, is helpful in some contexts, but it is of no use here.

terproductive, unnecessarily consuming resources and possibly resulting in unwarranted exclusion of probative evidence.

The major arguments for and against heightened scrutiny of scientific evidence were rehearsed earlier.¹⁷⁵ The principal problem is not that it is difficult for lay factfinders to assess an expert's reasoning or conclusions without possessing the underlying expertise. That much is true of all expert testimony. If there is a rationale for a special rule for scientific experts, it must be something special about science that justifies stricter scrutiny. Three features of scientific expert testimony provide this rationale: (1) Science is generally more difficult to understand than other areas of expertise;¹⁷⁶ (2) science is not only relatively impenetrable, but it is more impressive than non-scientific evidence, posing a special danger that jurors will give too much weight to evidence that carries with it the trappings of scientific truth;¹⁷⁷ and (3) until a period of rigorous testing passes, few scientists will be available to testify to the limitations or risks of errors in a scientific analysis. As a result, the usual safeguards of the trial process—cross-examination and opposing testimony—may be unavailable or ineffective.

With these reasons for an especially demanding screening of scientific evidence in mind, the boundary problem becomes tractable. The court should consider whether these three concerns are pre-

¹⁷⁵ See *supra* Section I.B.1.

¹⁷⁶ Some opinion polling research suggests that a large fraction of the U.S. populace does not appreciate basic scientific ideas. See, e.g., Paul Recer, *Most Americans Support Scientific Research, Poll Finds*, *Portland Oregonian*, July 2, 1998, at A13 (stating that a survey conducted for the National Science Foundation found that 61% of U.S. adults thought lasers work by focusing sound waves, 49% thought that humans lived at the same time as dinosaurs, 52% thought that the earth orbits the sun in one day or one month rather than one year, and 89% were unable to define a molecule); Rebecca Zacks, *What Are They Thinking?*, *Sci. Am.*, Oct. 1997, at 34 (giving the results of surveys and interviews with 1,200 freshmen at ten colleges indicating that approximately 45% reject the theory of evolution). Other polls show that people form beliefs on the basis of evidence that most scientists would find unconvincing. E.g., *Poll: U.S. Hiding Knowledge of Aliens*, CNN Interactive, at <http://www.cnn.com/US/9706/15/ufo.poll/index.html> (June 15, 1997) (reporting that 64% of Americans believe that creatures from elsewhere in the universe have contacted human beings, and 50% believe that aliens have abducted humans); Richard Ruelas, *10% in State Look to Skies*, *See UFOs*, *Ariz. Republic*, July 26, 1997, at A1 (reporting that 9% of people polled in Maricopa County, Arizona's most populous and urbanized region, say they have seen what they "really believe to be spaceships from other planets in the Arizona sky").

¹⁷⁷ See *supra* note 27.

sent in sufficient degree to warrant heightened scrutiny. Under this approach to the boundary problem, mathematical modeling of physical or biological processes such as the flow of water¹⁷⁸ or the survival of wildlife,¹⁷⁹ applications of mathematical equations that yield computer enhancement of images,¹⁸⁰ or statistical or econometric modeling of many types of data¹⁸¹ might seem to qualify for heightened scrutiny.¹⁸² Although these methods do not involve sophisticated laboratory instruments, they can be inscrutable and impressive to the uninitiated. It is not easy to shrug off a "best fit" or a "maximum likelihood estimate." Indeed, as we have seen, the California Supreme Court once was so moved by a trivial and inadequately countered bit of mathematics as to brand mathematics "a veritable sorcerer in our computerized society."¹⁸³

Nevertheless, it is not clear that *Frye* or *Daubert* (or some variant) should be applied to particular forms of mathematical and statistical modeling. Unlike a new chemical test or a novel physical theory or instrument, which might require significant time and experimental effort to probe, the adequacy, limits, or untested assumptions of most mathematical and statistical models can be defined fairly readily by other experts. Consequently, effective opposing testimony is generally available (if the economics of the

¹⁷⁸ See, e.g., *United States v. Sargent County Water Res. Dist.*, 876 F. Supp. 1090, 1095 (D.N.D. 1994) (using the "HEC-2" computer model to show the flow of water through a drain); *Am. Water Dev., Inc. v. City of Alamosa*, 874 P.2d 352, 368 (Colo. 1994) (en banc) (questioning the credibility of competing models of geologic and hydrologic features).

¹⁷⁹ E.g., *Seattle Audubon Soc'y v. Lyons*, 871 F. Supp. 1291, 1320-21 (W.D. Wash. 1994) (discussing northern spotted owl population dynamics), *aff'd*, 80 F.3d 1401 (1996).

¹⁸⁰ Cf. *United States v. Quinn*, 18 F.3d 1461, 1464-65 (9th Cir. 1994) (admitting "photogrammetry" testimony about computer-assisted calculations involving a rate of change in the sizes of receding objects as compared to objects of known size, over defense request for a hearing on the soundness of the technique).

¹⁸¹ See, e.g., *Statistical Methods in Discrimination Litigation*, *supra* note 14; Rubinfeld, *supra* note 19, at 200-03 (suggesting guidelines for courts to use when evaluating the qualifications of, and methodologies used by, experts in regression analysis); Daniel L. Rubinfeld, *Econometrics in the Courtroom*, 85 *Colum. L. Rev.* 1048, 1049 (1985) (arguing that "instead of accepting statistical rules of thumb such as the five percent significance test, courts should use an instrumentalist, efficiency-oriented criterion for determining appropriate standards of proof").

¹⁸² Commentators have assumed as much without discussion. See, e.g., *Mueller & Kirkpatrick*, *supra* note 27, § 7.8, at 991-92.

¹⁸³ *People v. Collins*, 438 P.2d 33, 33 (Cal. 1968).

case warrant it). It is unlikely that jurors will be overwhelmed with one side's set of equations when the other side can produce another set of equations or results. Indeed, triers of fact sometimes seem as ready to embrace fallacious criticisms of models as to recognize valid objections to them. Thus, condition (3) does not hold, and the import of condition (2) is unclear in this context.

In the end, however, it is condition (1) that should be decisive; statistical studies should not be exempt from careful scrutiny under standards like general acceptance or scientific soundness. As with Gresham's Law, bad statistical proof drives out (or at least devalues) the good.¹⁸⁴ The perception that statistics can prove anything, combined with the typical aversion to mathematics, makes it all too easy for quite dubious statistical analyses to appear the equal of far sounder assessments.¹⁸⁵ If the end result of a liberal policy of admissibility is the proverbial battle of the experts, with jurors no better able to decide the case when the fighting ceases, then the cost of the campaign is a dead-weight loss.¹⁸⁶ For these reasons, complex statistical testimony warrants some level of heightened scrutiny.

In the discussion that follows, I consider how the scrutiny required under *Daubert* and *Frye* should be applied to such studies. In federal jurisdictions, however, the Supreme Court's decision in

¹⁸⁴ Named for Sir Thomas Gresham (1519–79), the law states that “if two coins are in circulation whose relative face values differ from their relative bullion content, the ‘dearer’ coin will be extracted from circulation for melting down.” Hence, the law often is abbreviated to “bad money drives out good.” Graham Bannock et al., *The Penguin Dictionary of Economics* (1998).

¹⁸⁵ See, e.g., *United States v. Hall*, 165 F.3d 1095, 1119 (7th Cir. 1999) (Easterbrook, J., concurring) (“Delivering a graduate level statistical-methods course to jurors is impractical, yet without it a barrage of expert testimony may leave the jurors more befuddled than enlightened. Many lawyers think that experts neutralize each other, leaving the jurors where they were before the process began. Many lawyers think that the best (most persuasive) experts are those who have taken acting lessons and have deep voices, rather than those who have done the best research.”).

¹⁸⁶ It might be thought that a similar loss arises when two valid ways to look at the data lead to opposite conclusions. However, highly probative evidence should not be excluded because the opposing party also has strong evidence. For example, credible witnesses often have different recollections of the same conversation. Unless the evidence is in the range where its probative value is substantially outweighed by its prejudicial effect or other counterweights to its admission—a judgment that can sometimes be made with specialized, categorical rules in lieu of ad hoc balancing—a litigant must be permitted to introduce the evidence. Otherwise, no close case could go to trial.

*Kumho Tire Co. v. Carmichael*¹⁸⁷ relieves the pressure to define a clear boundary between science and nonscience. There, the Court wrote that all expert testimony must meet the "reliability" standard announced in *Daubert*, but that not all the factors used to ascertain scientific validity might apply, or they might apply differently to other areas of expertise. *Kumho* arose in response to a fatal automobile accident caused by a tire failure. The district court excluded an engineer's testimony that a manufacturing defect led to a separation between the tire tread and an internal structure, known as a steel-belted carcass, causing a blowout. This court applied the standard for scientific evidence described in *Daubert* to find that the engineer's analysis of his "visual inspection" of the tire lacked a sound "scientific basis."¹⁸⁸ The Court of Appeals for the Eleventh Circuit reversed the resulting summary judgment on the theory that "a *Daubert* analysis applies only where an expert relies on the application of scientific principles, rather than on skill- or experience-based observation."¹⁸⁹

In an opinion written by Justice Stephen G. Breyer, the Supreme Court reversed the court of appeals and held that the district court's exclusion of the engineer's analysis was not an abuse of discretion.¹⁹⁰ Every Justice agreed that Federal Rule 702 means that a witness testifying as an expert must present expert "knowledge"¹⁹¹ rather than speculation, and that "where such testimony's factual basis, data, principles, methods, or their application are called sufficiently into question, . . . the trial judge must determine whether the testimony has 'a reliable basis in the knowledge and experience

¹⁸⁷ 526 U.S. 137 (1999).

¹⁸⁸ *Id.* at 146.

¹⁸⁹ *Id.* (internal quotations omitted) (quoting *Carmichael v. Samyang Tire, Inc.*, 131 F.3d 1433, 1436 (11th Cir. 1997) rev'd, *Kumho Tire Co. v. Carmichael*, 526 U.S. 137 (1999)).

¹⁹⁰ Only Justice Stevens dissented, and even he joined most of the majority opinion. He would have remanded the case to the court of appeals to decide whether the trial court had abused its discretion under the principles outlined in the majority opinion. *Id.* at 159 (Stevens, J., concurring in part and dissenting in part).

¹⁹¹ *Id.* at 147. The rule stated that: "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." *Id.* (quoting Fed. R. Evid. 702 (prior to 2000 amendment)).

of [the relevant] discipline.”¹⁹² Finally, the Court wrote that, in making the determination that the expert was providing specialized knowledge that was sound enough to assist the trier of fact, the trial judge “may consider [the] more specific factors [enumerated in] *Daubert*.”¹⁹³

In short, *Kumho* extends *Daubert*’s call for “evidentiary reliability” and “a valid . . . connection to the pertinent inquiry as a precondition to admissibility[]”¹⁹⁴ to all expert testimony, but it discerns no universal solvent for ascertaining the validity of putative expert knowledge.¹⁹⁵ Some assurance of validity is required even from “experts in drug terms, handwriting analysis, criminal modus operandi, land valuation, agricultural practices, railroad procedures, attorney’s fee valuation, and others,”¹⁹⁶ but in such situations the details of *Daubert* may not apply,¹⁹⁷ and it is unclear what *Kumho* demands.¹⁹⁸ When it comes to engineering analysis that “rests upon scientific foundations,”¹⁹⁹ however, *Kumho* strongly suggests that the central considerations articulated in *Daubert*—the extent to which a theory or technique has been tested and subjected to critical scientific inquiry—are vital.²⁰⁰

¹⁹² *Id.* at 149 (alteration in original) (quoting *Daubert*, 509 U.S. at 592).

¹⁹³ *Id.* at 149–50 (emphasis omitted) (quoting *Daubert*, 509 U.S. at 592).

¹⁹⁴ *Id.* at 149 (omission in original) (internal quotations omitted) (quoting *Daubert*, 509 U.S. at 592).

¹⁹⁵ The Court states that:

[T]he trial judge must have considerable leeway in deciding in a particular case how to go about determining whether particular expert testimony is reliable. That is to say, a trial court should consider the specific factors identified in *Daubert* where they are reasonable measures of the reliability of expert testimony.

Id. at 152.

¹⁹⁶ *Id.* at 150 (emphasis omitted).

¹⁹⁷ *Id.*

¹⁹⁸ The three concurring Justices who also joined the majority opinion cautioned that although “the *Daubert* factors are not holy writ, in a particular case the failure to apply one or another of them may be unreasonable, and hence an abuse of discretion.” *Id.* at 159 (Scalia, O’Connor, and Thomas, JJ., concurring). But which cases are these? How can we tell whether real expertise exists unless the theories, techniques, and their practitioners have been subject to meaningful testing? For discussion of these questions, see Joseph Sanders, *Kumho* and How We Know, 64 *Law & Contemp. Probs.* 373 (2001).

¹⁹⁹ *Kumho*, 526 U.S. at 150.

²⁰⁰ The *Kumho* Court intimates that trial judges should be fairly demanding, as was the district court in *Kumho* itself. The Court pointedly observes that:

The same principle should govern the use of statistical methods. The statistical theory or technique should be one that has been subjected to sufficient study to establish its validity as applied to a class of problems that includes the one being investigated in the litigation.²⁰¹ Whether such a method is being applied properly to the problem at hand is a separate question that the Supreme Court, regrettably, has conflated with the issue of the validity of the method itself.²⁰² I turn now to that topic.

2. *The Usurpation Problem and the Methodology-Conclusion Puzzle*

Before *Daubert*, it was clear that the elevated scrutiny reserved for scientific evidence applied to the methodology that an expert employed rather than the conclusions that the expert reached by applying that methodology to specific facts. When heightened scrutiny is confined to methodology, the usurpation problem is manageable. Jurors are free to accept or reject particular conclusions as long as they are derived with an acceptable methodology and not otherwise subject to exclusion.²⁰³ In *Frye v. United*

[S]ome of *Daubert's* questions can help to evaluate the reliability even of experience-based testimony. In certain cases, it will be appropriate for the trial judge to ask, for example, how often an engineering expert's experience-based methodology has produced erroneous results, or whether such a method is generally accepted in the relevant engineering community.

Id. at 151.

²⁰¹ Cf. Richard O. Lempert, *The Jury and Scientific Evidence*, 9 Kan. J.L. & Pub. Pol'y 22, 23 (1999) ("At least so long as the expert's field is one requiring technical knowledge of a type that might be validated by science (compare a tire expert with, for example, an expert on fly fishing), the judge's role should be the same [as in *Daubert*].").

²⁰² It has been argued that the Court's treatment of the evidence in *Kumho* suffers from this flaw:

[W]hen I first read the trial judge's decision in *Kumho Tire* and the Court of Appeals' decision that reversed the trial judge, I thought that this was a "junk science" case, and it had been correctly decided. But after reading the briefs from both sides, looking for what seemed to be the likely facts, I began to think that the plaintiff's evidence in *Kumho Tire* was not "junk science" at all. It turns out that the methods used by the plaintiff's expert were the same as those used by the defendant's expert; they just reached different conclusions.

Id. at 26.

²⁰³ Committing the threshold question of the adequacy of the methods to judges rather than juries is a sensible procedural response to the need to screen out inadequately validated scientific evidence. See Edward J. Imwinkelried, *Judge Versus Jury: Who Should Decide Questions of Preliminary Facts Conditioning the*

States,²⁰⁴ for example, the court of appeals spoke of “testimony deduced from a well-recognized scientific principle or discovery”²⁰⁵ and the need to ensure that “the thing from which the deduction is made [has been] sufficiently established to have gained general acceptance in the particular field in which it belongs.”²⁰⁶ The court upheld the exclusion of the psychologist’s testimony, not because of doubts about how well he conducted the test on the defendant, but because “the systolic blood pressure deception test has not yet gained such standing and scientific recognition.”²⁰⁷ If the expert’s reasoning were recast in syllogistic form,²⁰⁸ it might proceed along the following lines:

Major Premise P₁: All subjects whose systolic blood pressure remains constant as they answer questions about their alleged participation in crimes are answering truthfully.

Minor Premise P₂: The systolic blood pressure of Alphonse Frye, who was accused of a crime, remained constant as he asserted his innocence in answering questions about the murder.

Conclusion C: Frye was telling the truth when he denied committing the murder.

Only the major premise *P₁* is subject to general acceptance “among physiological and psychological authorities.”²⁰⁹ The minor premise *P₂*, which is specific to the case, is more like the testimony of any other witness about his or her observations. It is not an expression of esoteric scientific reasoning, and it would make little sense to ask whether the scientific community generally accepts a case-specific proposition such as the particular blood pressure

Admissibility of Scientific Evidence?, 25 Wm. & Mary L. Rev. 577 (1984); Imwinkelried, *supra* note 7.

²⁰⁴ 293 F. 1013 (D.C. Cir. 1923).

²⁰⁵ *Id.* at 1014.

²⁰⁶ *Id.*

²⁰⁷ *Id.*

²⁰⁸ The syllogistic formulation is offered for heuristic purposes. Cf. Hand, *supra* note 22, at 51 (describing the expert’s role in supplying “the major premise”). The reasoning is inductive, not deductive. See Brian Skyrms, *Choice and Chance: An Introduction to Inductive Logic* 6–7 (3d ed. 1986).

²⁰⁹ *Frye*, 293 F. at 1014; cf. *Glastetter v. Novartis Pharms. Corp.*, 252 F.3d 986, 989 (8th Cir. 2001) (“Although this chain of medical reasoning appears sound, its major premise remains unproven. Glastetter’s experts failed to produce scientifically convincing evidence that Parlodel causes vasoconstriction.”).

readings taken from a single individual. Ordinary procedures like cross-examination can test whether the witness is speaking truthfully when he testifies that defendant's blood pressure did not rise.²¹⁰

In some respects, this dichotomy between the major and minor premises is oversimplified to bring out the methodology-conclusion distinction as sharply as possible.²¹¹ The complications, however, do not affect the basic point. Indeed, they help enucleate the principle that underlies the distinction between conclusion and methodology. Among other things, a full analysis would recognize that, in addition to deducing *C* (that Frye was telling the truth), Marston deduced the minor premise P_2 from another logical argument about the sphygmograph used to chart Frye's blood pressure. That argument might have as its major premise a claim P_1' that the instrument Marston used was capable of recording systolic blood pressure accurately. The minor premise P_2' of the supplemental argument would relate to the measurements that Marston made on Frye himself. The general acceptance test would apply to this additional major premise P_1' about the ability of the instrument to measure blood pressure, but not to the case-specific minor premise P_2' about the sphygmogram obtained in this particular case. The latter could be tested by having an opposing expert explain how the recording could have erroneously reflected the true blood pressure curve or by cross-examination to this effect. By definition, case-specific facts are not subject to "general acceptance" but must be determined on a case-by-case basis.

The basic point, then, is that whenever an expert's chain of reasoning includes general propositions that cut across cases and that

²¹⁰ Of course, the mere fact that a witness can be cross-examined is not a reason to admit the testimony. But, if the evaluation of cross-examination (or conflicting testimony) would require the jury to evaluate the soundness of the science that the witness relies on—a task that the jury is not well suited to undertake—then there is a reason for heightened scrutiny.

²¹¹ Professor Edward J. Imwinkelried has argued convincingly that this dichotomy has illuminated the distinction between the aspects of an expert's testimony that are subject to Federal Rule 702 and those that are governed by Federal Rule 703. See Edward J. Imwinkelried, *The "Bases" of Expert Testimony: The Syllogistic Structure of Scientific Testimony*, 67 N.C. L. Rev. 1, 5 (1989) [hereinafter Imwinkelried, *The "Bases" of Expert Testimony*]; Edward J. Imwinkelried, *The Educational Significance of the Syllogistic Structure of Expert Testimony*, 87 Nw. U. L. Rev. 1148, 1149–52 (1993) [hereinafter Imwinkelried, *Educational Significance*].

are purportedly scientific, these claims—and only these claims—should be subject to special scrutiny. The crucial distinction, in other words, is between the case-specific facts asserted in minor premisses and the trans-case facts asserted in major premisses.²¹² The former are “adjudicative facts,” while the latter are “legislative facts.”²¹³ Screening for general acceptance prevents the jury from relying on a legislative fact—the validity of a scientific theory—when the fact is not generally accepted in the relevant community of experts.

Daubert works no change in the principle, clearly established under *Frye*, that the heightened scrutiny pertains strictly to methodology.²¹⁴ Instead, *Daubert* simply substitutes for the pure general-

²¹² Cf. Imwinkelried, *The “Bases” of Expert Testimony*, supra note 211 (same); Edward J. Imwinkelried, *Developing a Coherent Theory of the Structure of Federal Rule of Evidence 703*, 47 Mercer L. Rev. 447, 476, 478 (1996) (arguing that this distinction is important in defining the scope of Federal Rule 703); Imwinkelried, *Educational Significance*, supra note 211 (same); Edward J. Imwinkelried, *The Meaning of “Facts or Data” in Federal Rule of Evidence 703: The Significance of the Supreme Court’s Decision to Rely on Federal Rule 702 in Daubert v. Merrell Dow Pharms., Inc.*, 54 Md. L. Rev. 352, 358–59 (1995) (same).

²¹³ The terminology was coined by Professor Kenneth Davis. 2 Kenneth Culp Davis, *Administrative Law Treatise* § 15.03, at 353 (1958) (defining “adjudicative facts” as those “concerning the immediate parties — who did what, where, when, how, and with what motive or intent”); Kenneth Culp Davis, *An Approach to Problems of Evidence in the Administrative Process*, 55 Harv. L. Rev. 364, 402, 408–09 (1942) (defining “legislative facts” as those involved “[w]hen an agency wrestles with a question of law or policy . . . the [more general] facts which inform its legislative judgment”). Professors John Monahan and Laurens Walker have built on this basic distinction in considering ways to inform factfinders of relevant social science knowledge. John Monahan & Laurens Walker, *Social Authority: Obtaining, Evaluating, and Establishing Social Science in Law*, 134 U. Pa. L. Rev. 477 (1986).

²¹⁴ In an early case on the admissibility of DNA testing, the trial court in *People v. Castro*, 545 N.Y.S.2d 985, 995 (Sup. Ct. 1989), grafted a “third prong” on to the normal requirements that the underlying theory and method of reaching expert conclusions be generally accepted. This court insisted on “a pre-trial hearing to determine if the testing laboratory performed the accepted scientific techniques in analyzing the forensic samples in this particular case.” *Id.* Routinely treating the application of a generally accepted theory and methodology as an aspect of admissibility represents a novel extension of *Frye*. The *Castro* court justified this additional “prong” on the theory that:

Given the complexity of the DNA multi-system identification tests and the powerful impact that they may have on a jury, passing muster under *Frye* alone is insufficient to place this type of evidence before a jury without a preliminary, critical examination of the actual testing procedures performed in a particular case.

acceptance test a richer set of criteria with which to scrutinize methodology. Under both *Daubert* and *Frye*, “[t]he focus, of course, must be solely on principles and methodology, not on the conclusions that they generate.”²¹⁵

In *Daubert* itself, this focus became quite blurred. The excluded testimony was the experts’ opinion that Bendectin was a human teratogen. Was the underlying “methodology” the unpublished reanalysis of data from a published epidemiological study, as the Ninth Circuit had thought? Was it the undisclosed statistical procedure used in this reanalysis to discern a statistically significant association between exposure to Bendectin and limb reduction defects? Was it inferring teratogenicity in humans in the absence of consistent and statistically significant epidemiological findings? Or, is it possible that the experts’ opinion was itself a “methodology” that required a preliminary showing of soundness? The Supreme Court’s discussion of scientific soundness was so abstract and unconnected to the evidence in the case that its opinion provides no answer. On remand, the Ninth Circuit also gave no answer, and it refused to let the district court venture into this thicket. Rather, it upheld the summary judgment on the ground that even if general causation could be proved, the admissible evidence could not support a conclusion that the plaintiffs’ injuries were attributable to the drug.²¹⁶

The analysis offered above provides some answers to this inquiry. In *Frye*, the case-specific conclusion was that the defendant was telling the truth when he denied being the murderer (C). In

Id. at 987. Some courts accepted this innovation. See, e.g., *United States v. Martinez*, 3 F.3d 1191, 1197–98 (8th Cir. 1993); *Ex parte Perry*, 586 So. 2d 242, 248–50 (Ala. 1991). Most *Frye* jurisdictions are satisfied with pretrial hearings on general acceptance alone. See, e.g., *State v. Vandebogart*, 616 A.2d 483, 495 (N.H. 1992); *State v. Cauthron*, 846 P.2d 502, 507 & n.4 (Wash. 1993).

²¹⁵ *Daubert*, 509 U.S. at 595.

²¹⁶ See *Daubert v. Merrell Dow Pharms., Inc.*, 43 F.3d 1311, 1320–21 (9th Cir. 1995) (emphasis and citations omitted). The court reasoned that:

California tort law requires plaintiffs to show not merely that Bendectin increased the likelihood of injury, but that it more likely than not caused their injuries. . . . In terms of statistical proof, this means that plaintiffs must establish not just that their mothers’ ingestion of Bendectin increased somewhat the likelihood of birth defects, but that it more than doubled it—only then can it be said that Bendectin is more likely than not the source of their injury.

Id. Here, however, “[n]one of plaintiffs’ epidemiological experts claims that ingestion of Bendectin during pregnancy more than doubles the risk of birth defects.” *Id.*

Daubert, the analogous case-specific conclusion is that Bendectin caused plaintiffs' injuries (C"). These are adjudicative facts in the two cases. The methodology-conclusion distinction focuses attention at the stage of admissibility on the legislative facts—the scientifically established, trans-case premises used in reaching the case-specific conclusions. In *Daubert*, these premises include the proposition that Bendectin is a teratogen—that it can (and sometimes does) cause limb reduction defects (P₁"). Thus, if a single expert had been offered to prove C" (specific causation), the gate-keeping role would have required elevated scrutiny of the underlying scientific premise P₁" (general causation).

The plaintiffs divided up the reasoning from the various premises to the case-specific conclusion C" among several experts, however. One group was willing to attest to general causation, and a different group to specific causation. This division of expert labor can make no difference in applying the methodology-conclusion distinction. *Daubert* is simply a case in which one expert's testimony ends at the methodological level of the major premise, and another expert's testimony employs that premise to reach the case-specific conclusion.²¹⁷ It is comparable to having one expert testify that a sudden systolic pressure spike is indicative of deception and another report that, because he found no such spike, defendant was not deceptive. Under *Frye*, the first expert's "conclusion" about the physiological correlate of deception would have to be generally accepted. The second expert's case-specific observations would not have to run this gauntlet. Under *Daubert*, the only dif-

²¹⁷ Schematically, the proof might look something like the following:

P₁" : Sufficient exposure in utero to Bendectin can cause injury. (Expert 1)

P₂" : Plaintiff received sufficient exposure. (Lay witness and Expert 2)

P₃" : Plaintiff was injured. (Expert 3)

P₄" : Plaintiff was not exposed to anything else that could have injured him. (Expert 3)

P₅" : The injury did not occur spontaneously. (Expert 3)

C" : Bendectin caused plaintiff's injury. (Expert 3)

Regardless of how the testifying is divided up, P₁" is a legislative fact to which heightened scrutiny should be applied. P₂" through P₅" are adjudicative facts, but all of them could (and some clearly do) rest on unstated scientific premises that warrant scrutiny. Under *Daubert*, such scrutiny is required of all the scientific premises of the argument. That the judge may find that some of the premises are only weakly established, and that the conclusion is in grave doubt, goes to the sufficiency of the evidence to support the conclusion C". These flaws in the proof are not themselves a ground for exclusion.

ference is that the first expert's "conclusion" would have to be adequately validated by reference to general acceptance *and* other factors.

Recognizing that the labels "methodology" and "conclusion" can be confusing, and lacking a well articulated standard for using these terms, courts in recent years have shied away from them. *General Electric Co. v. Joiner*²¹⁸ is the most prominent example. Robert Joiner was an electrician who worked for nearly twenty years for a city water and light department in Georgia. His work brought him into contact with Polychlorinated biphenyls ("PCBs") in electrical transformers. In 1991, at the age of thirty-seven, he was diagnosed with lung cancer.²¹⁹ Joiner and his wife sued three manufacturers of PCBs on theories of strict liability, negligence, and fraud.²²⁰ A former cigarette smoker, Joiner alleged that tobacco smoke acted as an initiator of his cancer and that the PCBs acted as a promotor, transforming the initiated cells into malignant growths.²²¹ Defendants moved for summary judgment.²²² They argued that "plaintiffs . . . cannot present credible, admissible scientific evidence that . . . small cell lung cancer in humans can be caused or promoted by PCBs,"²²³ and they maintained that PCBs do not cause cancer unless other chemicals—namely, furans or dioxins—are present. Plaintiffs' experts pointed to studies of PCBs to dispute this claim,²²⁴ and they suggested that there were reasons to think that Joiner had been exposed to PCBs, furans, and dioxins. Defendants argued further, however, that the available evidence indicated that Joiner had no significant exposure to any of these three types of chemicals.²²⁵

The district court granted the defendant's motion for summary judgment. It found that although there was a genuine dispute as to

²¹⁸ 522 U.S. 136 (1997).

²¹⁹ *Joiner v. Gen. Elec. Co.*, 864 F. Supp. 1310, 1312–13 (N.D. Ga. 1994), rev'd, 78 F.3d 524 (11th Cir. 1996), rev'd, 522 U.S. 136 (1997).

²²⁰ *Id.* at 1314.

²²¹ *Id.* at 1313–14.

²²² The cases were initiated in state court, but defendants removed them to the United States District Court for the Northern District of Georgia. *Id.* at 1314.

²²³ *Id.* (omissions in original) (internal quotations omitted).

²²⁴ *Id.* at 1322–27 (discussing studies of PCBs).

²²⁵ *Joiner v. Gen. Elec. Co.*, 78 F.3d 524, 528 (11th Cir. 1996), rev'd, 522 U.S. 136 (1997).

whether Joiner was exposed to PCBs, the potentially admissible evidence failed to show that he was exposed to furans or dioxins.²²⁶ Furthermore, the court found that the epidemiological and animal studies on which plaintiffs' experts relied were too weak to justify the conclusion that PCBs can promote cancers. Finding this inajor premise scientifically unsound, the district court ruled the expert testimony that rested on it to be inadmissible.

A divided panel of the Eleventh Circuit reversed. Two judges concluded that the district court "improperly assessed the admissibility of the proffered scientific expert testimony and overlooked evidence establishing disputed issues of fact."²²⁷ In particular, the court held that there was a disputed issue of fact as to whether Joiner was exposed to furans and dioxins, and that the district court erred in finding the claim that PCBs promote cancers to be too speculative to be admissible.

The Supreme Court granted certiorari to review the "particularly stringent standard of review"²²⁸ that the court of appeals purported to apply to the district court's ruling that plaintiffs' experts' opinions were inadmissible under *Daubert*. The Supreme Court unanimously agreed that the district court's ruling on admissibility was reversible only for an abuse of discretion,²²⁹ and all but one Justice²³⁰ agreed that the district court's ruling excluding the experts' opinions about the effects of PCBs was within its discretion.²³¹ The portion of the majority opinion upholding the

²²⁶ *Joiner*, 864 F. Supp. at 1318-19.

²²⁷ *Joiner*, 78 F.3d at 528 (opinion of Judge Barkett). Judge Birch "concur[red] in this opinion," emphasizing in his concurring opinion that the district court exceeded its role as "gatekeeper" by examining the weight and sufficiency of the evidence. *Id.* at 534 (Birch, J., concurring).

²²⁸ *Id.* at 529.

²²⁹ This result was predictable, but it has been criticized. Compare D.H. Kaye, *Joiner* and *Scheffer*: Scientific Evidence in the Supreme Court, Newsletter (Ass'n of Am. Law Sch. Sec. on Evidence), Fall 1997, at 2, with Michael J. Saks, *The Aftermath of Daubert: An Evolving Jurisprudence of Expert Evidence*, 40 *Jurimetrics J.* 229, 235-36 (2000).

²³⁰ *Joiner*, 522 U.S. at 152 (1997) (Stevens, J., concurring in part and dissenting in part).

²³¹ The Supreme Court remanded the case so that the district court could address what it perceived as two unresolved questions: "Whether Joiner was exposed to furans and dioxins, and whether if there was such exposure, the opinions of Joiner's experts would then be admissible." *Id.* at 147. The district court, however, already had concluded that even "[a]ssuming that Plaintiff's experts had not made unfounded

evidentiary ruling reviewed the research literature on whether PCBs promote cancers and concluded that the district court did not err in finding that the experts could not establish this major premise in a scientifically sound manner.²³²

This disposition required the Court to confront the argument "that because the District Court's disagreement was with the conclusion that the experts drew from the studies, the District Court committed legal error and was properly reversed by the Court of Appeals."²³³ According to Justice John Paul Stevens:

The reliability ruling was more complex and arguably is not faithful to the statement in *Daubert* that "[t]he focus, of course, must be solely on principles and methodology, not on the conclusions that they generate." Joiner's experts used a "weight of the evidence" methodology to assess whether Joiner's exposure to transformer fluids promoted his lung cancer. They did not suggest that any one study provided adequate support for their conclusions, but instead relied on all the studies taken together (along with their interviews of Joiner and their review of his medical records). The District Court, however, examined the studies one by one and concluded that none was sufficient to show a link between PCB's and lung cancer. The focus of the opinion was on the separate studies and the conclusions of the experts, not on the experts' methodology.

Unlike the District Court, the Court of Appeals expressly decided that a "weight of the evidence" methodology was scientifically acceptable.²³⁴

Rather than analyze the methodology-conclusion distinction, the majority threw up its hands:

assumptions about [exposure to] furans and dioxins, . . . Plaintiffs' expert testimony would not be admissible." *Joiner*, 864 F. Supp. at 1322.

²³² See *Joiner*, 522 U.S. at 145-47. A group of distinguished scientists had argued that plaintiffs' experts' views on PCBs were unfounded. Their brief reviewed the epidemiological literature and concluded that "there were no human studies showing any causal relationship between PCBs and small-cell lung cancer . . ." Brief of Amici Curiae Bruce N. Ames et al. in Support of Petitioners at 20, *Joiner*, 78 F.3d 524 (No. 96-188). These amici also insisted that "[t]he mice studies . . . did not fit the facts of this case." *Id.* at 19-20. The Court's opinion reiterated these conclusions.

²³³ *Joiner*, 522 U.S. at 146.

²³⁴ *Id.* at 152-53 (Stevens, J., concurring in part and dissenting in part) (citations omitted).

Respondent points to *Daubert's* language that the "focus, of course, must be solely on principles and methodology, not on the conclusions that they generate." . . . But conclusions and methodology are not entirely distinct from one another. Trained experts commonly extrapolate from existing data. But nothing in either *Daubert* or the Federal Rules of Evidence requires a district court to admit opinion evidence that is connected to existing data only by the *ipse dixit* of the expert. A court may conclude that there is simply too great an analytical gap between the data and the opinion proffered.²³⁵

This abandonment of the focus on methodology prompted Justice Stevens to retort:

Daubert quite clearly forbids trial judges to assess the validity or strength of an expert's scientific conclusions, which is a matter for the jury. Because I am persuaded that the difference between methodology and conclusions is just as categorical as the distinction between means and ends, I do not think the statement that "conclusions and methodology are not entirely distinct from one another," either is accurate or helps us answer the difficult admissibility question presented by this record.²³⁶

As Justice Stevens maintained, the distinction between methodology and conclusion is viable,²³⁷ but the classification serves legal rather than scientific purposes and must be applied accordingly. The words function to avoid excessive scrutiny of case-specific, minor premises and case-specific conclusions. The trans-case, major premise that PCBs promote cancers in human beings should be shown to be sufficiently well-established by the methods of science to justify its use in an expert chain of reasoning.²³⁸ The majority's

²³⁵ *Id.* at 146 (citation omitted).

²³⁶ *Id.* at 154-55 (Stevens, J., concurring in part and dissenting in part) (citations omitted).

²³⁷ See Saks, *supra* note 229, at 235-36.

²³⁸ Consequently, both the majority and the dissent in *Joiner* are correct. The dissenting opinion is correct to criticize the majority for eliding the fundamental distinction, but it is wrong to treat "weight of evidence" as a valid "methodology" that scientists employ (when they write review papers or develop risk assessments). *Joiner*, 522 U.S. at 152-53 (Stevens, J., concurring in part and dissenting in part). Validity at so abstract a level is not enough to ensure that the putative scientific theories that reach juries are sufficiently scientifically sound to justify a jury's reliance on testimony about them (or derived from them). The majority is correct to demand more than the *ipse dixit* of an expert that the literature supports the major premise, but wrong to

demand that the expert not leap to a conclusion about the carcinogenicity of PCBs²³⁹ is consistent with this specificity analysis.

Following *Joiner*, however, the Supreme Court has continued to blur the methodology-conclusion distinction. In *Kumho Tire Co. v. Carmichael*, the Court observed that “[t]he objective of [*Daubert*] is to . . . make certain that an expert, whether basing testimony upon professional studies or personal experience, employs in the courtroom the same level of intellectual rigor that characterizes the practice of an expert in the relevant field.”²⁴⁰ Various lower federal courts had drawn the same lesson from *Daubert*, and several have spoken of a departure from the level of professional care normally observed outside of litigation as a reason to exclude statistical testimony.²⁴¹ Because the *Kumho* opinion deals with all stripes of experts, including those who rely on skill that is not reducible to any articulated methodology, the search for some substitute for the “scientific methodology” standard sketched in *Daubert*²⁴² is understandable and important.

Kumho’s quasi-malpractice standard is useful in this connection, but the demand for ordinary rigor should not excuse the failure of an entire field of putative experts to apply truly rigorous standards in developing their field. Neither should it result in the exclusion of expert testimony just because a judge believes that a more rigorous analysis would have led to different conclusions. A demand for “rigor” is easy to apply to all facets of expert testimony—

contend that this is because “conclusions and methodology are not entirely distinct from one another.” *Id.* at 146.

²³⁹ *Id.* (“A court may conclude that there is simply too great an analytical gap between the data and the opinion proffered.”).

²⁴⁰ *Kumho*, 526 U.S. at 152.

²⁴¹ See, e.g., *Sheehan v. Daily Racing Form, Inc.*, 104 F.3d 940, 942 (7th Cir. 1997) (describing the omission of certain data and variables as “a failure to exercise the degree of care that a statistician would use in his scientific work, outside of the context of litigation,” and citing *Braun v. Lorillard Inc.*, 84 F.3d 230, 234 (7th Cir. 1996), *Rosen v. Ciba-Geigy Corp.*, 78 F.3d 316, 318 (7th Cir. 1996), and *Daubert v. Merrell Dow Pharms., Inc.*, 43 F.3d 1311, 1316–19 (9th Cir. 1995), for the proposition that “*Daubert* . . . requires the district judge to satisfy himself that the expert is being as careful as he would be in his regular professional work outside his paid litigation consulting”).

²⁴² For elaborations on the opinion’s efforts to describe the process of evaluating scientific theories, see Foster & Huber, *supra* note 172; Erica Beecher-Monas, A Ray of Light for Judges Blinded by Science: Triers of Science and Intellectual Due Process, 33 Ga. L. Rev. 1047, 1051–55 (2000).

conclusions as well as methods. It could tempt courts to exclude legitimately debatable testimony that they find unpersuasive, even though it is based on generally accepted and valid methods. To be sure, there will be cases in which an expert has been so sloppy in applying these methods that the testimony would not be sufficiently probative under Federal Rule 403, but the stricter scrutiny reserved for trans-case scientific reasoning should not be applied under the rubric of rigor to case-specific conclusions.²⁴³

In sum, the specificity standard for distinguishing methodology from conclusion for the purpose of applying heightened scrutiny is superior to the *Joiner* Court's apparent willingness to allow the category of methodology to bleed into the category of conclusions. It is also superior to any tendency to read into *Kumho* a requirement that case-specific conclusions be subjected to the careful scrutiny that is properly reserved for scientific methods.

Nonetheless, the specificity standard is not always trivial to apply. In particular, problems can arise in screening statistical evidence, which typically involves methods that are accepted at a very general level and that are sound as applied to certain types of data but not to others. For example, whether an expert has used an acceptable formula for estimating the frequency of a genotype in the population plainly is a methodological issue. It involves a trans-case, major premise. Equally plainly, whether the same expert has done the arithmetic correctly is a case-specific question not subject to heightened scrutiny under *Frye* or *Daubert*. But consider *State v. Garcia*,²⁴⁴ in which:

[A] trial court in Arizona admitted testimony about likelihood ratios in a rape case involving two assailants. . . . [A]nalysis of the semen stain on the victim's blouse indicated that sperm from two males were present. According to the court of appeals, a population geneticist "provided the jury with likelihood ratios (broken down by population subgroups such as Cauca-

²⁴³ The amended Federal Rule 702 requires that "the witness has applied the principles and methods reliably to the facts of the case." Fed. R. Evid. 702(3). This condition implements *Joiner's* willingness to examine an expert's conclusions as an aspect of admissibility. If applied too strictly, it will exclude probative expert evidence that is like almost all evidence—it has weaknesses, but jurors can appreciate this fact and give the testimony the weight it deserves. Thus, this prong of the amended rule should be applied in a manner that is consistent with Federal Rule 403.

²⁴⁴ 3 P.3d 999 (Ariz. Ct. App. 1999).

sians, African Americans, and the like) for three distinct scenarios involving the sources of the DNA mixture found in the stain: (1) victim, defendant and unknown versus victim and two unknowns; (2) victim, defendant and unknown versus defendant and two unknowns; and (3) victim, defendant and one unknown versus three unknowns.²⁴⁵

The trial court admitted this testimony following a *Frye* hearing at which the state's expert testified to general acceptance. The defendant was convicted. On appeal, he argued that the state had not proved that the specific formulas used to calculate the likelihood ratios had been generally accepted. The court of appeals affirmed the conviction, reasoning that both the concept of the likelihood ratio and the specific formulas were generally accepted, as indicated by publications in the scientific literature.

In a petition for review, Garcia suggested that although the use of the likelihood ratio has support in the literature, the particular formulas were not previously published. There is no general formula, however, for computing a likelihood ratio. The formula depends on the specific hypotheses being compared. The likelihood ratio for a mixture with two possible men is different from that for a mixture with three, or four, and so on. The same approach produces the appropriate expression in each situation, and arriving at the correct expression is like solving word problems in high school algebra. Everyone agrees that the problems should be solved with formulas derived according to the rules of algebra, but different word problems require different formulas. The use of algebra is generally accepted, but a student can make a mistake applying those rules.

In *Garcia*, the use of likelihood ratios is generally accepted as scientifically valid, but an expert can make a mistake in algebraically representing the pertinent conditional probabilities or in working out the algebra that yields the likelihood ratio for a par-

²⁴⁵ 1 Modern Scientific Evidence, *supra* note 23, § 15-5.4, at 251-52 (Supp. 2000) (quoting *State v. Garcia*, 305 Ariz. Adv. Rep. 7, ¶ 15 (Ct. App. 1999)). To illustrate the nature of the testimony with simplified numbers for the first set of hypotheses, the calculations might show that the chance of the specified DNA types being present was 100 times greater if (a) the DNA came from the victim, the defendant, and a randomly selected person than if (b) it came from the victim and two randomly selected persons.

ticular problem. Is this a concern about the case-specific, minor premise (so that *Frye* would not apply) or a trans-case, major premise (that must be generally accepted)? Because the formulas used in *Garcia* easily could be employed in other cases involving a mixture of DNA from one female and two males, they fall into the latter category. There would be little difficulty admitting them under *Daubert*, for the derivation of the formulas is a straightforward algebraic exercise that can be verified by any number of experts familiar with probability theory.²⁴⁶ Affidavits from a few such experts should be enough to demonstrate the requisite reliability. Under *Frye*, it is more difficult to introduce even an obviously valid result that has yet to be scrutinized fully by the relevant portion of the scientific community, but an advocate can build a record of acceptance even in this situation.²⁴⁷ In any event, the added difficulty of satisfying *Frye* is not a reason to depart from the specificity standard for the methodology-conclusion classification. If anything, it is a reason to replace *Frye* with a more direct inquiry into scientific validity.

D. Looking Back at Statistical Evidence

The prior sections reveal that until recently, statistical evidence either was admitted as a matter of course, excluded as irrelevant because it was obviously baseless, or questioned on extremely du-

²⁴⁶ Comparison could be made to a great many results in applied mathematics. For instance, the MacLaurin series for expressing any differentiable function as an infinite series is well known. See, e.g., The Chemical Rubber Co., Standard Mathematical Tables 407-08 (Samuel M. Selby ed., 14th ed. 1965). The method is generally accepted, but it produces different expressions for different functions. For example, the function e^{ax} becomes $1 + ax + a^2x^2/2! + a^3x^3/3! + \dots$, while the function $\sin(bx)$ becomes $bx - b^3x^3/3! + b^5x^5/5! - b^7x^7/7! + \dots$. Very similar expansions can be found in standard references such as the CRC tables, and any mathematician can check whether these expansions are correct.

Whether the population genetics models that give rise to some of the expressions that enter into a likelihood ratio for DNA mixtures are sufficiently validated is a distinct question from the accuracy of the algebra. The adequacy of these models in a given situation is ultimately an empirical question rather than a mathematical one.

²⁴⁷ For instance, in *Andrews v. State*, 533 So. 2d 841 (Fla. Dist. Ct. App. 1988), Lifecodes Corporation had a prominent molecular biologist from MIT visit its laboratory and testify that the DNA techniques that Lifecodes used to identify the defendant as a rapist and burglar were generally accepted, notwithstanding the novelty of their application to establishing human identity. Id. at 847-49.

bious grounds. In each of these instances, the strict scrutiny standards for scientific evidence were not applied to statistical proof. As late as 1994, it could be said that although "a particular study may use a method that is . . . so poorly executed that it should be inadmissible[.]. . . [m]ore often . . . the battle over statistical evidence concerns weight or sufficiency rather than admissibility."²⁴⁸ Indeed, the 1999 edition of *McCormick on Evidence* does not even address the subtleties of applying the special standards for scientific evidence to statistical analyses, for it suggests that the admissibility of statistical assessments rarely is in doubt.²⁴⁹ With the explosion of employment discrimination claims brought under Title VII of the 1964 Civil Rights Act in the 1970s and 1980s, and through the efforts of economists²⁵⁰ and statisticians in a broad spectrum of cases, courts became exposed to—and came to expect²⁵¹—more sophisticated and potentially more useful statistical models.²⁵² To be sure, there was no shortage of argument among experts and counsel about the persuasiveness of specific statistical analyses.²⁵³ Many courts experienced considerable difficulty pene-

²⁴⁸ David H. Kaye & David A. Freedman, Reference Guide on Statistics, *in* Reference Manual on Scientific Evidence 333, 335 n.2 (Federal Judicial Center ed., 1994).

²⁴⁹ See 1 McCormick on Evidence, *supra* note 25, § 209, at 800.

²⁵⁰ See, e.g., Arnold H. Lozowick et al., Law and Quantitative Multivariate Analysis: An Encounter, 66 Mich. L. Rev. 1641 (1968); Rubinfeld & Steiner, *supra* note 18.

²⁵¹ E.g., Thornburg v. Gingles, 478 U.S. 30, 53 n.20 (1986) (describing how the court turned to the literature to assess the methods of analysis used); Coble v. Hot Springs Sch. Dist. No. 6, 682 F.2d 721, 730–33 (8th Cir. 1982) (chiding plaintiffs for not applying multiple regression analysis).

²⁵² By 1984, a legal news reporter could observe that "[w]hat demonstrative evidence was to the 1960s and early '70s, statistics have become to the 1980s—the hottest new way to prove a complicated case." David Lauter, Making a Case with Statistics, *Nat'l L.J.*, Dec. 10, 1984, at 1.

²⁵³ See, e.g., McCleskey v. Kemp, 753 F.2d 877, 888 (11th Cir. 1985) ("The usefulness of statistics obviously depends upon what is attempted to be proved by them."), *aff'd*, 481 U.S. 279 (1987); Chang v. Univ. of R.I., 606 F. Supp. 1161, 1206 (D.R.I. 1985) ("The studies by Siskin and Zellner reach diametrically opposed results with virtually the same data. But, there are numerous—and important—differences in the models."); *Presseisen v. Swarthmore Coll.*, 442 F. Supp. 593, 619 (E.D. Pa. 1977) ("It seems to the Court that each side has done a superior job in challenging the other's regression analysis, but only a mediocre job in supporting their own. In essence, they have destroyed each other and the Court is, in effect, left with nothing.").

trating these arguments,²⁵⁴ and some jurisdictions searched for bright-line rules that would reveal which statistical convention or procedure had to be used to produce a prima facie case.²⁵⁵ The admissibility of the studies, however, rarely was questioned.²⁵⁶

This situation changed as commentators and advocates brought concerns about “junk science” to the forefront of the judicial consciousness. Although *Daubert* was but a variation on the theme of earlier cases, the allusion to “gatekeeping” struck a responsive chord,²⁵⁷ encouraging federal district courts to be bolder in excluding scientific evidence and prompting state courts to reconsider their rules and to look more carefully at proffers of scientific testimony. Today, “*Daubert* motions” to exclude statistical studies or conclusions have migrated from the realm of epidemiology in which *Daubert* was grounded to many substantive fields and types of statistical proof. To identify the special issues that arise with sta-

²⁵⁴ See, e.g., *Wilkins v. Univ. of Houston*, 654 F.2d 388, 410 (5th Cir. 1981) (“In closing, we add a note both rueful and cautionary. The bar is reminded that sound statistical analysis is a task both complex and arduous. Indeed, obtaining sound results by these means, results that can withstand informed testing and sifting both as to method and result, is a mission of comparable difficulty to arriving at a correct diagnosis of disease.”), vacated and remanded, 459 U.S. 809 (1982); *Penk v. Or. State Bd. of Higher Educ.*, 48 Fair Empl. Prac. Cas. (BNA) 1724, at 1849, 1862 n.13 (D. Or. 1985) (noting that aspects of the regression analysis procedure are “technical in nature and difficult to grasp,” that the results may sometimes be “invalid and misleading,” and that “the methodology of this regression was not fully comprehensible to the court, and hence it is difficult to evaluate its results”), aff’d, 816 F.2d 458 (9th Cir. 1987).

²⁵⁵ See, e.g., *Palmer v. Shultz*, 815 F.2d 84, 92–95 (D.C. Cir. 1987) (holding that significance at 0.05 level in two-tail test required to create prima facie case). For criticism of the efforts to convert statistical practices to legal rules, see Mikel Aickin, *Issues and Methods in Discrimination Statistics*, in *Statistical Methods in Discrimination Litigation*, supra note 14, at 159; Laurens Walker & John Monahan, *Social Facts: Scientific Methodology as Legal Precedent*, 76 Cal. L. Rev. 877 (1988).

²⁵⁶ One reason may have been the lack of jury trials in Title VII and other civil cases that often involved multiple regressions. Excluding these studies is less critical in bench trials.

²⁵⁷ Justice Blackmun introduced the metaphor of federal judges as “gatekeepers” into the literature on scientific evidence in *Daubert*. See supra text accompanying note 159. The phrase has become so ubiquitous that there now are references to the “science” of gatekeeping. John M. Conley & David W. Peterson, *The Science of Gatekeeping: The Federal Judicial Center’s New Reference Manual on Scientific Evidence*, 74 N.C. L. Rev. 1183 (1996). The phrase has no special meaning. In applying the rules of evidence and procedure to exclude testimony, judges have been gatekeepers both before and after the adoption of evidence codes.

tistical expert testimony, and to illustrate how these issues should be approached, the remainder of this paper examines a study of damages in a major antitrust case.

II. BRINGING *DAUBERT* AND *KUMHO* TO BEAR: *CONWOOD CO. V. UNITED STATES TOBACCO CO.*²⁵⁸

A. *Conwood's Complaint: Monopolizing Moist Snuff*

Snuff is a smokeless tobacco product²⁵⁹ that is placed in small amounts between the cheek and the gums. The major producer of moist snuff is United States Tobacco Company, Inc. ("USTC"),²⁶⁰ followed by Conwood Company, L.P.²⁶¹ In 1998, Conwood filed a complaint in the United States District Court for the Western District of Kentucky alleging that USTC monopolized the moist snuff

²⁵⁸No. 00-6267 (6th Cir. Jan. 4, 2000). I must caution the reader that I received compensation for work on Appellants' briefs to the court of appeals and on the preparation for the oral argument.

²⁵⁹Smokeless tobacco products include loose leaf chewing tobacco, plug and twist chewing tobacco, dry snuff, and moist snuff.

The firms that have the largest market shares in the industry are U.S. Tobacco Company (37.9%), Conwood Company L.P. (23.2%), and Pinkerton Tobacco Company (28.1%). These firms together accounted for 83% of total U.S. production in the industry in 1996. Other firms in the industry include National Tobacco Company (9.2% market share), Swisher International Group, Inc. (6.8%), Brown & Williamson (0.5%), and R.C. Owen Company of Tennessee, Inc. (0.4%).

Edward Knight et al., *The U.S. Tobacco Industry in Domestic and World Markets*, Cong. Research Serv. Rep. No. 98-506 E, at CRS 21 (1998).

²⁶⁰U.S. Tobacco Company, Inc. is the holding company for United States Tobacco Company. Through subsidiaries, USTC manufactures and markets various consumer products and entertainment services. It is the world's leading producer of moist smokeless tobacco, with sales of 46 million pounds in 1996 and manufacturing facilities in Illinois, Kentucky and Tennessee. *Id.* USTC was created in the court-ordered dissolution of the Duke Tobacco Trust in 1911. Its brands account for approximately 75% of moist snuff sales in the United States. Proof Brief for Appellees at 5-6, *Conwood*, (No. 00-6267) [hereinafter Appellees' Brief].

²⁶¹"Conwood Company L.P. is a limited partnership which manufactures moist and dry snuff and loose leaf, plug and twist chewing tobacco. Estimated 1996 sales were 28 million pounds . . . [with] manufacturing facilities in Kentucky, North Carolina, and Tennessee." Knight et al., *supra* note 259, at CRS 210.

market in the U.S. in violation of Section 2 of the Sherman Act.²⁶² Conwood's theory, as developed at a four-week trial, was that:

In 1990, UST began an orchestrated campaign to choke off the distribution of rivals' products. Disdaining competition on the merits—which UST feared would erode its market share and profit margin—UST used its power to exclude competitors' display racks, advertising, and products. UST's representatives tossed as many as 20,000 Conwood [sales] racks [in retail stores] into dumpsters *each month*.²⁶³

USTC denied engaging in systematic, exclusionary conduct of this (or any other) sort. It moved to exclude econometric testimony designed to prove that USTC's allegedly illegal conduct gravely suppressed Conwood's sales of its brands of snuff, and it sought summary judgment. The district court denied these motions. At trial, USTC cross-examined Conwood's expert and presented its own expert who dismissed the damages study as worthless,²⁶⁴ but produced no evidence of its own as to the amount of damages.

After deliberating for under four hours, a jury awarded Conwood \$350 million in damages.²⁶⁵ Trebling this figure,²⁶⁶ the district court entered a judgment of \$1.05 billion.²⁶⁷ USTC's appeal to the Court of Appeals for the Sixth Circuit is pending.

B. Conwood's Resistance Theory

In establishing damages, Conwood relied heavily on an analysis prepared by Dr. Richard Leftwich, a professor of accounting and finance.²⁶⁸ As presented, the study appeared to be a paradigm of

²⁶² 15 U.S.C. § 2 (1997). Conwood also alleged various state law causes of actions but dropped these before the case went to the jury. Proof Brief for Appellants at 3, *Conwood*, (No. 00-6267) [hereinafter Appellants' Brief].

²⁶³ Appellees' Brief, *supra* note 260, at 2.

²⁶⁴ See *infra* note 294.

²⁶⁵ Appellees' Brief, *supra* note 260, at 4.

²⁶⁶ 15 U.S.C. § 15(a) (1997) (allowing a prevailing plaintiff to "recover threefold the damages by it sustained and the cost of suit").

²⁶⁷ Appellants' Brief, *supra* note 262, at 4.

²⁶⁸ Leftwich, who worked as a consultant with the firm of Lexecon, Inc., is "the Fuji Bank-Heller Professor of Accounting and Finance in the graduate school of business" at the University of Chicago. Leftwich Trial Transcript at 5, *Conwood*, (No. 00-6267) [hereinafter Leftwich Trial Transcript]. Professor Leftwich's academic training, as described in his curriculum vitae, is in accounting, finance, and applied economics,

objective, scientific inquiry. It began with a “test [of] a hypothesis about the effect of UST’s behavior on Conwood’s performance.”²⁶⁹ “The hypothesis was that UST’s anticompetitive behavior had a greater impact on Conwood’s market performance in cases where Conwood had a relevantly [sic] low market share in . . . 1990.”²⁷⁰ We can call this a “resistance theory.” Stated more fully, this theory posits that (1) USTC engaged in anticompetitive conduct to roughly the same degree in every state; (2) the conduct had little or no effect on Conwood’s sales in states where Conwood was resistant to these practices—where it had a large market share in 1990; and (3) the conduct had a greater effect on Conwood’s sales in states where Conwood was susceptible—where it had a small market presence in 1990.

C. Conwood’s Data

To test this resistance theory, the expert compiled a table (reproduced in the Appendix as Table A1) showing Conwood’s percentage of moist snuff sales in each state in 1990 and 1997.²⁷¹ For example, in 1990 Conwood sold 14% (by weight) of all moist snuff in Vermont; by 1997, Conwood’s share rose four percentage points to 18%. In the District of Columbia, the share rose 10.3 points, from 7.2% to 17.6%.²⁷² All these “raw data,” as Leftwich called them,²⁷³ are shown in Table A1 of the Appendix.

The figures in Table A1, however, are not those used in the expert’s first report. The original data came from an accounting

with a Ph.D. in applied economics and finance from the University of Rochester. Chicago GSB Faculty, Richard Leftwich, at <http://gsb.uchicago.edu> (last visited Oct. 1, 2001). At trial, he stated that he held “a Ph.D. in economics” and “an endowed chair” at a school “famous for . . . the economics department, the so-called Chicago School of Economics.” Leftwich Trial Transcript, *supra*, at 5–6.

The CEO of Conwood was allowed to opine as to the percentage of the market “lost” to USTC and to testify that each such point was worth ten million dollars in annual profits. Appellants’ Brief, *supra* note 262, at 50. Yet he was not designated as an expert under Fed. R. Civ. P. 26(a)(2), and even if he had been qualified as an expert, it is hard to imagine how the testimony could have survived a *Kumho* objection.

²⁶⁹ Leftwich Trial Transcript, *supra* note 268, at 11.

²⁷⁰ *Id.* at 12.

²⁷¹ The table omits data on sales in Hawaii and Alaska but includes the District of Columbia.

²⁷² The percentages are rounded off to the nearest tenth.

²⁷³ Leftwich Trial Transcript, *supra* note 268, at 17.

firm's report on the pounds of moist snuff sold annually in the various states.²⁷⁴ These numbers were not recorded correctly for the initial analysis. These data-entry errors resulted in an excess of \$245 million in estimated damages.²⁷⁵ Such errors are flaws in execution that should be evaluated under Federal Rule 403; they do not affect the validity of the statistical methodology. Under *Kumho*, it also could be argued that they bespeak a lack of rigor that should preclude the expert from testifying.²⁷⁶ Data-entry errors are common in academic research, however, and once the expert has corrected the major errors, even if belatedly, exclusion on this ground does not seem justified. A corrected analysis may well be based on a valid and (ultimately) a reasonably implemented approach.

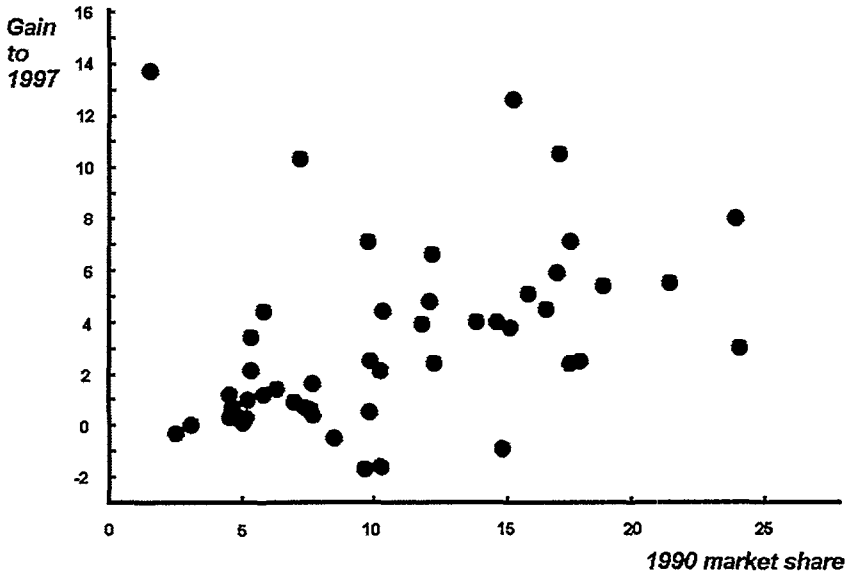
The state-by-state data can be presented more perspicaciously in graphical form. Figure 1 is a scatter diagram that plots the 1990 market share (the horizontal distance on the *X*-axis) against the subsequent growth (the height on the *Y*-axis). Each state thus appears as a point in the graph.

²⁷⁴ *Id.* at 47.

²⁷⁵ *Id.* at 45–46.

²⁷⁶ Indeed, the Court of Appeals for the Fifth Circuit recently invoked such reasoning in upholding the exclusion of a statistical analysis in a Title VII case. See *Munoz v. Orr*, 200 F.3d 291, 301 (5th Cir. 2000) (citing *Kumho* to hold that plaintiffs' statistical analysis was properly excluded as unreliable for problems ranging "from particular miscalculations to [the expert's] general approach to the analysis," including tables that did not add to anywhere near 100%).

Figure 1.
Scattergram for Conwood's Market Share Data



D. Regression Analysis to Show Causation

1. The Regression Results

Leftwich then testified that he could learn little by looking at the situation in particular states, for these results were just “anecdotes” or “stories.”²⁷⁷ “[A]s a professional economist,”²⁷⁸ he was obliged to undertake “systematic analyses” and “empirical data analysis.”²⁷⁹ Therefore, he used “a standard economic method . . . called regression analysis” to test “the prediction of the original hypothesis that Conwood’s performance in low market share states should have been . . . hampered more than it was in high market share states.”²⁸⁰ The “standard economic method” revealed that “there was a highly reliable relationship between Conwood’s growth in the pe-

²⁷⁷ Leftwich Trial Transcript, *supra* note 268, at 18.

²⁷⁸ *Id.* at 17.

²⁷⁹ *Id.* at 18.

²⁸⁰ *Id.*

riod [from 1990 to] 1997 and its market share in 1990.”²⁸¹ That is, “the results were highly reliable, or statistically significant; in . . . that there was more than a 95% chance . . . that these results were, in fact, reflective of systematic patterns in the data.”²⁸²

These characterizations of statistical significance and the nature of the relationship²⁸³ are misleading at best,²⁸⁴ but they result from a flawed attempt to translate technical terms into lay language²⁸⁵ and not necessarily from a failure to use sound statistical methods. As such, even though they do not reflect the “intellectual rigor” with which knowledgeable experts would be expected to present their results outside of litigation, they have no trans-case implications. Moreover, they can be fuel for effective impeachment. Therefore, these errors in the presentation of the statistical analysis should not preclude all testimony about the analysis.

The type of regression performed in *Conwood* is known as “simple linear regression.” The idea is to relate subsequent growth to initial market share with a straight line through the cloud of data points in Figure 1. The equation for a straight line that has a slope β and a height α where it intersects the Y -axis is

$$Y = \alpha + \beta X. \quad (1)$$

²⁸¹ Id. at 18–19.

²⁸² Id. at 20.

²⁸³ The assertion that the statistical relationship between the variables is “highly reliable” when, as explained below, that relationship “explains” only 13% of the variance in market share growth could have been used to impeach the witness’s understanding of the regression results or the care with which he approached the case. The assertion went unchallenged at trial, however.

²⁸⁴ They are instances of the “transposition fallacy.” See, e.g., Kaye & Freedman, supra note 15, at 131 & n.167. The 0.95 figure is not the probability in favor of the hypothesis that there is some association between initial market shares and subsequent growth. Rather, if certain assumptions hold and if there is no true association, then the probability is under 0.05 that the measured association would be as large as it was (or larger). This significance level, as it is called, suggests that there is some nonzero association, but neither the probability of this conclusion nor the extent of the association can be derived from the observed significance level.

²⁸⁵ The use of the term “confidence” for a significance level also is problematic. See, e.g., Aickin, supra note 255, at 170.

The Greek letters α and β stand for numbers that define a straight line,²⁸⁶ and the regression procedure simply finds the particular numbers that define the one line that best fits the data.²⁸⁷

Because factors apart from Conwood's shares in 1990 affect Conwood's market share in 1997, we would not expect the growth within each state in the 1990–97 period to be given exactly by this simple equation. Due to the many variables not captured in equation (1), in some states the growth will be greater, and in others it will be less. If the effects of all the unobserved factors merely combined to produce random fluctuations from the straight-line relationship, we could just add an "error term" to the equation to account for these disturbances. Conwood's expert therefore posited the following statistical model:

$$Y = \alpha + \beta X + \varepsilon, \quad (2)$$

where α is the growth expected in a state in which Conwood had no sales in 1990 (the Y -intercept), β is the constant increase in growth for a unit increase in initial market share, and ε is a random fluctuation from the values of Y expected on the basis of α and β alone. In other words, the error term ε represents "noise" that distorts the deterministic relationship of equation (1). Furthermore, the expert assumed that the level of the noise (from all the factors that actually determined sales but are omitted from equation (1)) was the same in every state and that it was what engineers call "white noise."²⁸⁸

²⁸⁶ For example, when $\alpha = 0$ and $\beta = 1$, the straight line is $Y = 0 + (1 \times X) = X$. This is the line that goes through the origin at an angle of forty-five degrees.

²⁸⁷ The "least-squares" regression used in *Conwood* treats the sum of the squared deviations from the fitted line as the measure of "best fit."

²⁸⁸ "White noise" refers to an error term that generates errors that are described by a "normal curve." Collectively, the assumptions of a normal distribution of independently distributed errors with the same variance at each value of X are known as the "normal error assumptions." See generally Rubinfeld, *supra* note 19, at 212–13 (describing techniques for determining the precision of regression results). The claim that the observed level of association was statistically significant depends on these assumptions, which Leftwich apparently never tested.

The failure to perform any regression diagnostics might well be cast as a methodological flaw. In opposition, it could be argued that the normal error assumptions are routine and that if they seem to be inapposite in a particular case, the opposing party could show that by analyzing the data appropriately. Whether a jury would follow these arguments about the presence and implications of departures from

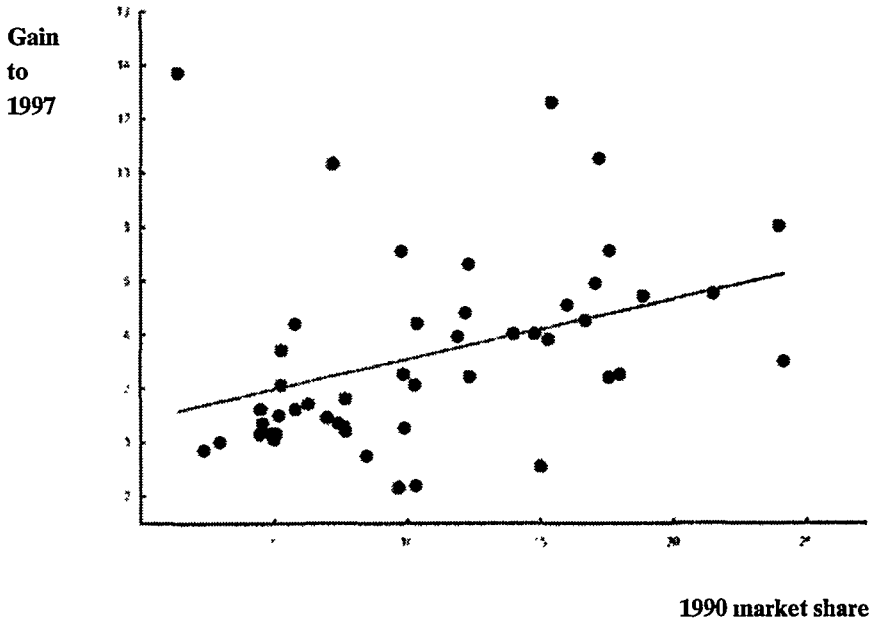
For the market share data of Table A1, the best estimate of the intercept α is 0.85, and the best estimate of the slope β is 0.22. That the estimated value of β is 0.22 means that, on average, across all states, every additional percentage point in the share of the 1990 market is associated with an increase of about two-tenths (0.22) of a percentage point by 1997.²⁸⁹ If there were no association at all ($\beta = 0$), and if other assumptions that Dr. Leftwich apparently did not verify held, then the chance that the observed value β would fall as far (or farther) from the expected value of zero as 0.22 would have been about 0.01. The regression line $Y = 0.85 + 0.22X$ is shown in Figure 2, which superimposes this straight line on the scattergram. Although the actual values show considerable dispersion about the estimated regression line, there is a modest correlation between Conwood's 1990 market share in a state and its subsequent share gain in that state.²⁹⁰

the normal error assumptions, however, is doubtful. The better approach is to place the burden of performing regression diagnostics on the proponent of the regression.

²⁸⁹ Doing a regression may not be the best way to test the resistance theory. Even if the exclusionary conduct is uniform or essentially random across states, the regression results depend on whether resistance varies smoothly with 1990 market share. A test that is less affected by the shape of the curve that relates resistance to 1990 market share is a simple comparison of means. For example, one can ask whether there is any difference in the average growth in states in which Conwood had less than 20% of the 1990 poundage than in states in which it had 20% or more. (Leftwich first used 20% as a cut-off point in computing damages. He also used 15%.) There is, but it is too small to be statistically significant (at the conventional 0.05 level). The same is true for the cut-point of 15%. Leftwich's testimony that "Conwood grew more in the foothold states, reliably more, in the foothold states than in the non-foothold states," Leftwich Trial Transcript, *supra* note 268, at 21-22, is difficult to comprehend.

²⁹⁰ The coefficient of variation, R^2 , is just over 13%. This number describes how well the straight line with this intercept and slope fits the data points in Figure 1. R^2 can range from 0.0 (for completely uncorrelated variables) to 1.0 (for perfectly correlated variables). Here, the estimated regression line "explains" only 13% of the total variance in market share growth among the states.

Figure 2.
The Regression Line



2. The Causal Inference

a. Applicability of Daubert

Even if there is a weak but real relationship between initial market share and subsequent growth, does it prove that “anticompetitive behavior hampered Conwood’s growth more in the non-toehold states than in the toehold states,”²⁹¹ as Conwood’s expert suggested? Or is the relationship, as USTC suggested on appeal, an exercise in searching for a pattern in noisy data and reading into that pattern something that is not there?²⁹²

²⁹¹ Leftwich Trial Transcript, *supra* note 268, at 30; see also *id.* at 10.

²⁹² Appellants’ Brief, *supra* note 262, at 55. Appellants argue that: Leftwich’s “low-share” theory has no foundation in economics or statistics. It never explored or explained what economic factors could give rise to this “low share” state pattern, but merely verified that the pattern conformed (with many exceptions) to that perceived by self-interested observers. That such patterns

The inference that the differences in Conwood's growth should be attributed to USTC's illegal acts requires a leap of faith, for the regression model contains no variable that measures these acts. One must step outside the regression framework to draw the desired conclusion, and this methodological step is difficult to justify under *Daubert*. The issue here is not just whether, under the facts of a specific case, certain assumptions in a statistical model are reasonable.²⁹³ The method in question requires the inference that illegal conduct caused injury to a competitor simply by positing some kind of resistance to illegal conduct that cannot be measured directly but, by hypothesis, is reflected in some pattern in the competitor's sales history after the conduct began. This logic could be used in any antitrust case. Being a general, seemingly scientific theory or procedure, the resistance theory should be subject to the full scrutiny that *Daubert* establishes for scientific evidence.

It is difficult to see how the resistance theory can survive this scrutiny. It has never been published or examined by other economists.²⁹⁴ As a procedure for discerning illegal conduct, the

exist is no surprise: "Almost any large data set — even pages from a table of random digits — will contain some unusual pattern . . ."

Id. (citations omitted).

²⁹³ In contrast, complaints about assumptions in a model might seem like an attack on the execution of a method rather than the validity of the method itself, but even this is not always so. Two assumptions implicit in the regression here already have been noted: (1) that the magnitude of exclusionary conduct in each state is the same, and (2) that the effective resistance increases linearly with initial market share. The failure to test these assumptions is a methodological flaw that goes to admissibility of the testimony that reaches the case-specific conclusion that USTC's conduct depressed Conwood's sales in the less resistant states. Nothing in economic theory or the specific data permits one to test the assumptions. The assumptions, if false, vitiate the inference of causation.

²⁹⁴ An economist hired by defendants for the litigation was critical of the damages study. David T. Scheffman Trial Transcript at 70, *Conwood* (No. 00-6267) ("[H]e's got a model that can't answer the question."); id. at 78 ("[H]e doesn't have a model that explains anything."); id. at 79-86 (describing other variables that should have been included in the regression model); id. at 86-87 ("[T]he data can't be used for what he used it [sic] for, and the model can't explain what he's trying to explain."). A courtroom crossfire between experts is not what *Daubert* contemplates when it speaks of peer review and scientific theories that have withstood efforts at falsification.

A second economist, Daniel L. McFadden, who was not involved in the trial, later reviewed the Leftwich study. Dr. McFadden is the Director of the Econometrics Laboratory, the E. Morris Cox Professor of Economics at the University of California, Berkeley, and the recipient of the 2000 Nobel Prize in Economic Sciences. He filed an amicus curiae brief in support of defendants-appellants. The brief states that it is submitted to avoid a result that "trivializes the important role that properly

resistance method could have an enormous error rate. The method is essentially circular. For example, an unscrupulous analyst intent on finding causation and damages could hypothesize that Conwood's marketing efforts are more susceptible to USTC's conduct in the mountain states of Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, and Wyoming.²⁹⁵ The analyst then could "confirm" this hypothesis with a critical ratio test on the data in Table A1, for Conwood's mean gain in market share in the mountain states is one-quarter of its gain in the states outside this mountain region.²⁹⁶

b. Implications of the Threat of Confounding

A further obstacle to inferring causation is the threat of confounding. "Confounding" refers to the action of an unobserved variable that also is correlated with the dependent and the independent variables.²⁹⁷ Without data on potential confounders, it is impossible to disentangle the effect of the measured variable from the potentially confounding ones. In the *Conwood* case, it is easy to suggest possible confounding variables. Perhaps personal income among snuff users has grown more in states in which Conwood had

conducted economic analyses can and should play in litigated matters." Brief of Amicus Curiae Dr. Daniel L. McFadden in Support of Defendants-Appellants at 1, *Conwood* (No. 00-6267) [hereinafter McFadden's Brief]. It concludes that "Dr. Leftwich's analysis . . . is fundamentally flawed," "cannot be relied upon," and "does not meet the standard for Daubert." *Id.* at 23. However, appellants objected to the court's considering the amicus brief, and the court of appeals ruled that it would not accept this brief.

²⁹⁵ This is the designation for these states used by the Census Bureau. See, e.g., Census Bureau, *State Population Estimates: Annual Time Series (1999)*, available at <http://www.census.gov>.

²⁹⁶ The mean gain in Conwood's market share in mountain states is 0.9 percentage points; in non-mountain states it is 3.7 percentage points. The difference is statistically significant at the 0.05 level.

²⁹⁷ *Kaye & Freedman*, *supra* note 15, at 92 n.24, 138. In an age-discrimination case, for instance, it may well be that age is correlated with lay offs—older workers are laid off at a higher rate than younger ones. This could mean that the older workers are laid off because they are older, or it could mean that they are laid off at a greater rate because they are less productive. If productivity correlates with age, then it is impossible to tell simply from the correlation between age and lay offs whether the workers are being laid off because they are old or because they are less productive. See, e.g., *Sheehan v. Daily Racing Form, Inc.*, 104 F.3d 940, 942 (7th Cir. 1997) (referring to "the more than remote possibility that age was correlated with a legitimate job-related qualification, such as familiarity with computers" in a job that required computer skills).

small shares in 1990, and USTC's brands appeal more to relatively affluent users. Population migration across state lines might be at work. Regional differences in consumer attitudes might lead to more growth in states in some regions than in others. Advertising restrictions and the conduct of other competitors also vary across states.

The well-known fact that correlation is not causation,²⁹⁸ however, is not itself a reason to exclude an observational study offered to prove causation.²⁹⁹ The validity of an inference of causation depends on how well the study succeeds in "controlling" for plausible confounders, and the extent to which its conclusions have been replicated in other populations.³⁰⁰ The most effective procedure for controlling for lurking variables is a randomized, controlled experiment.³⁰¹ Of course, that is not possible in most econometric research, and it was not possible in *Conwood*. With adequate data, however, a statistical analyst can determine whether another variable might account for the pattern. The analyst could "control" for income, for instance, by examining whether Conwood's share growth in those states where snuff users experienced similar income growth was related to Conwood's initial market share. Another approach would be to modify equation (2) by adding a variable for personal income growth among snuff users. If we call

²⁹⁸ See, e.g., *Glastetter v. Novartis Pharms. Corp.*, No. 00-3087, 00-3467, 2001 WL 630651, at *3 (8th Cir. 2001) ("Though case reports demonstrate a temporal association between Parlodel and stroke, or stroke-precursors, that association is not scientifically valid proof of causation."); *Tagatz v. Marquette Univ.*, 861 F.2d 1040, 1044 (7th Cir. 1988) (observing that although the statistical analysis indicates a correlation and shows that discrimination may be a possible cause, the statistics do not exclude other possible explanations).

²⁹⁹ But see *Munoz v. Orr*, 200 F.3d 291, 301 (5th Cir. 2000) (suggesting that under *Kumho*, it was proper to exclude plaintiffs' expert's statistical study in part because the expert "stated that discrimination was the 'cause' of the disparities he had observed, a statement which he later recanted as 'overzealous' since statistics can show only correlation and not causation").

³⁰⁰ See generally *Kaye & Freedman*, supra note 15, at 96 (discussing the application of a study's findings to population groups not involved in the study); *Zeisel & Kaye*, supra note 16, at 27-43 (discussing the inference of causation from observational data).

³⁰¹ Oddly, both professors described regression as a kind of experiment. See *Leftwich Trial Transcript*, supra note 268, at 19 (describing an agricultural experiment to test the effect of fertilizer on crop yield); *McFadden's Brief*, supra note 294, at 10, 23 (describing regression as a "scientific experiment").

this variable Z and use the Greek letter gamma (γ) to denote the change in market share (Y) associated with a unit change in Z (for a fixed value of the starting market share, X), then equation (2) becomes

$$Y = \alpha + \beta X + \gamma Z + \varepsilon. \quad (3)$$

If Z is correlated with X , then the estimated value of β should decline (relative to equation (2)), making it harder to attribute a change in market share growth (Y) to a unit change in initial market share (X).

Conwood's expert examined some possibly confounding variables with a multiple regression model similar to equation (3). He did not report whether they were correlated with 1990 market share (X), but stated that "I tested all of the plausible explanations that I had data that enabled me to test"³⁰² and that "[m]y tests showed that plausible alternative explanations were inconsistent with the patterns I found in the data."³⁰³ If the expert actually employed reasonable procedures to eliminate all plausible rival hypotheses, then the resistance-regression procedure should not be inadmissible simply because the initial regression left open the possibility of confounding variables.

c. Resistance Versus Momentum

One "plausible explanation" that the expert purported to eliminate was not a confounding variable, but rather went to the core of the resistance theory. Instead of attributing the change in market share to the hypothetical "resistance" to USTC's conduct in some states but not others, one might well suppose that there would be more growth, on average, in states where Conwood was better established, if only because its products, for any number of reasons, were selling better in those states than in others. In other words, the regression depicts the effects of "momentum" as readily as "resistance."

Conwood's expert purported to refute the momentum interpretation of the regression of 1990–97 growth on 1990 shares by means

³⁰² Leftwich Trial Transcript, *supra* note 268, at 22.

³⁰³ *Id.* at 30.

of a regression of 1984–90 growth on 1984 shares.³⁰⁴ This regression did not reveal any statistically significant association. Having already found a statistically significant association in the post-1990 period, he concluded that the only thing that could explain the change from “not significant” before 1990 to “significant” after 1990 was differential resistance to illegal conduct.

This reasoning is fallacious. The momentum theory asserts that with or without resistance, initial market share (X) tends to predict subsequent market growth (Y). A large change in the predictive value of the initial market share, as between the earlier and later periods, would undercut the theory that only momentum is at work in both periods (as opposed to momentum alone in the earlier period and momentum plus resistance to illegal conduct in the later period). At first glance, it looks like the change in the impact of initial market share in the pre-1990 period to the impact in the post-1990 period is substantial. The pre-1990 estimate of the slope is -0.13 , but the post-1990 estimate is 0.22 . Both these numbers, however, are estimates of the unknown slope β in equation (2). The true value of β in the pre-1990 period could be higher, and the true value in the post-1990 period could be lower. Before concluding that the difference in the two periods should be attributed to resistance to illegal conduct after 1990, the hypothesis that β is actually the same in both periods (and the observed difference is attributable to chance) must be rejected.³⁰⁵ Yet, Conwood’s expert never tested this hypothesis. Had he done so, he would have found that the uncertainty in the difference in the estimated slopes for each period is too large to permit the conclusion that the difference is statistically significant.³⁰⁶

³⁰⁴ *Id.* at 27.

³⁰⁵ McFadden’s Brief, *supra* note 294, at 13 (“The mere fact that the two coefficients are different, or that one is statistically significant while the other is not, does not constitute an appropriate statistical test of the hypothesis that the relationship . . . changed . . .”).

³⁰⁶ The difference between the estimated slopes in the two periods is

$$-0.13 - 0.22 = -0.35.$$

The standard error of this difference is

$$\sqrt{0.181^2 + 0.083^2} = 0.199.$$

The t-statistic is therefore

$$-0.35/0.199 = -1.75.$$

Although the point may seem a fine one, the failure of Conwood's expert to test for the significance of the *difference* in the estimates of the slope is a methodological flaw that affects the validity of his effort to refute the momentum theory. To make the same point with other language from *Daubert*, one can observe that the use of two separate tests for significance, rather than a single test of the difference between the two estimates, does not "fit" the problem of eliminating the rival momentum theory as an explanation for the pattern in the 1990–97 period.

E. Regression Analysis to Estimate Damages

1. Estimating Effect with a "Regression Rectangle"

If the resistance-regression proof of causation is vulnerable to assault under *Daubert*, the use of the regression analysis to estimate damages is open to mayhem. Conwood's expert treated USTC as the cause—indeed, the sole cause—of Conwood's lower growth in most states. As explained in the preceding section, he purported to verify this treatment by a statistical regression model that assumed that the market share in 1997 is equal to one constant plus a second constant multiplied by the market share that Conwood had in 1990.³⁰⁷ This regression did not take into account any variables to show the effect of USTC's alleged anticompetitive practices. It did not adequately consider whether the pattern or trend in market share growth changed before and after the time that the practices that were supposed to have depressed Conwood's growth were instituted.

For forty-nine data points, a difference of ± 1.75 standard errors (or more) has a probability of about 0.09 of occurring when β is the same in both periods. Because this p-value exceeds 0.05, the difference in the estimated slopes is not statistically significant at the conventional 0.05 level that Conwood's expert used elsewhere in his testimony.

³⁰⁷ Leftwich regressed gain in market share from 1990 to 1997 on 1990 share:

$$S_{1997} - S_{1990} = a + bS_{1990}$$

Adding S_{1990} to each side of the equation gives

$$S_{1997} = a + (b + 1)S_{1990}$$

In other words, examining the correlation between 1990 shares and 1990–97 gains is essentially the same as examining the correlation between 1990 shares and 1997 shares.

In computing damages, the expert inexplicably modified the actual market shares in a way that was supposed to account for the extent of USTC's "bad acts."³⁰⁸ As in the causation analysis, the 1990-97 growth (as adjusted) was regressed on 1990 shares, yielding the straight line $Y = 1.8 + 0.31X$, which is plotted in Figure 3.

Thus far, there has been no analysis of damages—just another regression showing a weak correlation between two variables. To arrive at a figure for damages, Conwood's expert divided the forty-nine states into two groups. The high-share, supposedly resistant group consisted of three states in which Conwood had more than 20% of the market in 1990. Although the law allows substantial latitude in estimating damages once liability has been established, the expert had no economic theory or data that indicated why he selected this cut-off point. Nevertheless, he assumed that these states were unaffected by USTC's anticompetitive practices and hence had no damages.

The low-share, susceptible group consisted of the other forty-six states. Leftwich assumed that had there been no anticompetitive practices, the 1997 Conwood market share in every one of these forty-six low-share states would have gone up from 1990 by the same amount. But he did not use the actual experience of the high-share states in the 20+% range to deduce this amount.³⁰⁹ Instead, he used the regression of 1997 on 1990 shares to predict that if Conwood started with 20% of a state's market in 1990, it would have 28.1% of the market in 1997. If Conwood's share in a low share state had gone up less than 8.1 percentage points, he boosted its

³⁰⁸ Leftwich did not collect data on or analyze the incidence of alleged "bad acts." However, USTC's economist did, and he performed a regression with individual stores as the unit of analysis that showed no effect. Leftwich used these data on "bad acts" assembled by USTC to modify the annual state market shares used in his regression study. Leftwich Trial Transcript, *supra* note 268, at 34. Plaintiffs' Exhibit No. 327.1 is a scatter diagram showing adjusted shares. Table A2 of the Appendix gives corresponding numbers. The use of these adjusted figures is inconsistent with the theory behind using "resistance" as an indicator of illegal conduct. The underlying theory is that differences between share growth in states with high and low resistance to USTC's conduct reflect that conduct. But if the effect of the conduct already is reflected in the market shares, then "resistance" cannot serve as an indirect indication of the suppression of market share resulting from USTC's conduct.

³⁰⁹ Had he done so, he would have had to report that Conwood's average growth in those 3 states was 5.5 percentage points, which is not statistically significantly different from the average gain of 3.0 percentage points in the 46 low-share states.

gain to 8.1; if Conwood's share had gone up more than 8.1 percentage points, he reduced the gain to 8.1 points. Thus, he gave every one of the low-share states a market gain of 8.1 points—an amount that exceeded Conwood's actual performance in two of the three high-share states that supposedly were unaffected by USTC's practices.

Figure 3.
Conwood's Estimate of How Much More of the Market it Would Have Gained

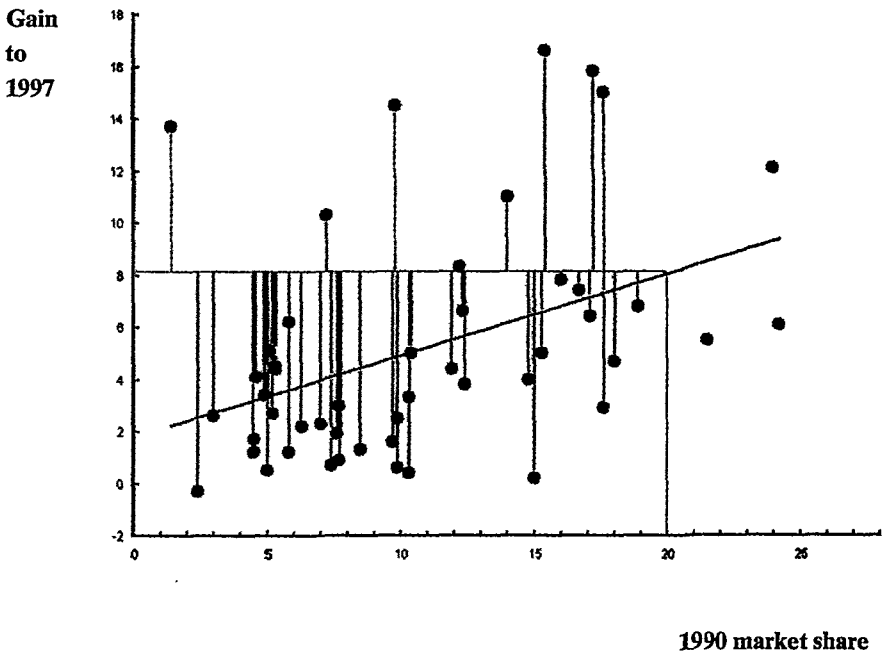


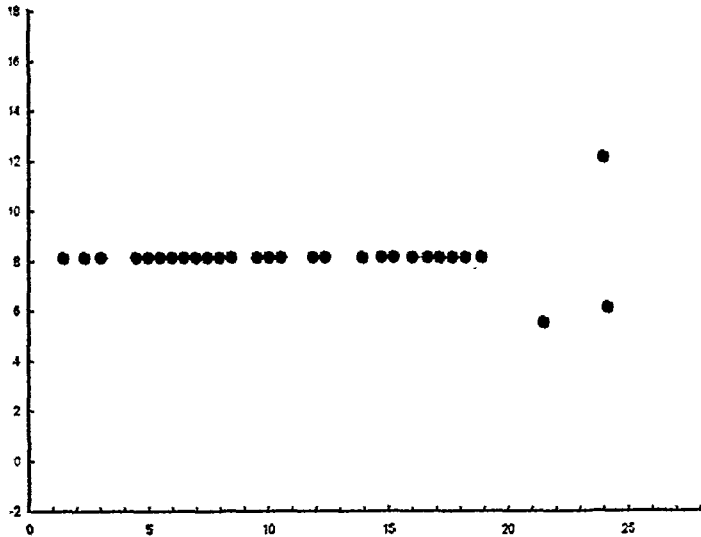
Figure 3 is a picture of this augmentation of market shares. The points in the rectangle are states in which Conwood supposedly would have gained more market share in the absence of USTC's acts. The lengths of the vertical lines drawn from these points up to the horizontal line $Y = 8.1$ are the increases in market share growth that Conwood's expert awarded Conwood in these states. The points above the rectangle are the states in which Conwood outperformed the gain expected of a state in which Conwood had 20%

of the market in 1990. The lengths of vertical lines drawn from these points down to the horizontal line $Y = 8.1$ are the decreases in market share growth given to these states. The net adjustment is the difference between the sum of the lengths of the first set of lines (in the rectangle) and the sum of those in the second set (above the rectangle). This difference translated into \$488 million dollars of estimated damages.³¹⁰

Figure 4.

Conwood's Estimate of the Market Shares It Would Have Gained Without "Bad Acts" by USTC

Gain to
1997



1990 market share

³¹⁰ Leftwich Trial Transcript, *supra* note 268, at 42. Leftwich performed a similar analysis using 15% as the cutoff for high-share states. The height of the rectangle then is only 6.5 (because the height of the regression line at $X = 15$ is only 6.5), and the width of the rectangle is only 15 (leaving 13 states in the resistant group that, *ex hypothesi*, suffered no damages). Based on this schema, Leftwich produced \$313 million as the lower bound for damages. Both the upper and the lower estimates ignore the statistical uncertainty in the predicted share values at the cut points of 15% and 20%; they also ignore possible errors in the data.

The expert's picture of what Conwood's growth would have looked like in the absence of the "bad acts" from 1990-97 is shown in Figure 4. The forty-six states in or above the rectangle now have the same share gain of 8.1.³¹¹

2. Applying Daubert to the "Regression Rectangle"

Skepticism of expert testimony is one thing; exclusion of that testimony is another. To apply the validity requirement of *Daubert* to this procedure for estimating damages, a court must ask whether the methodology is sound. Conwood argued to the district court that "regression analysis and other such economic models are accepted and tested methods for proving damages,"³¹² and the district court was satisfied with this rejoinder. In a brief opinion that recited the sources of the data and the fact that Conwood's expert applied various regressions to establish his resistance theory, the district court concluded that "Leftwich's testimony satisfies *Daubert*. His methodologies are generally acceptable. Defendant's expert also used them. . . . The credibility of the expert and his opinions is an issue for the jury."³¹³

By "methodologies," the court apparently meant the statistical procedure of regression. The court relied exclusively on another district court opinion, *Ohio v. Louis Trauth Dairy, Inc.*,³¹⁴ a price-fixing case that it characterized as holding that *Daubert* was satisfied because "the experts[] all . . . were economists or statisticians [who] conducted econometric and regression analyses that were testable, generally acceptable, and reproducible."³¹⁵ As I explain

³¹¹ Because the positions of many states on the line $Y = 8.1$ overlap, fewer than 49 points are visible in the figure.

³¹² Plaintiffs' Opposition to Motion to Exclude the Damages Study and Future Testimony of Dr. Richard Leftwich at 22, *Conwood Co. v. United States Tobacco Co.* (W.D. Ky. Dec. 9, 1999) (Civil Action No. 5:98CV-108-R).

³¹³ Memorandum Opinion and Order at 4, *Conwood Co. v. United States Tobacco Co.* (W.D. Ky. Feb. 23, 2000) (Civil Action No. 5:98CV-108-R) (citation omitted) [hereinafter Memorandum and Opinion Order].

³¹⁴ 925 F. Supp. 1247 (S.D. Ohio 1996).

³¹⁵ Memorandum Opinion and Order, *supra* note 313, at 4. The *Trauth Dairy* court stated: "Econometric and regression analyses are generally considered reliable disciplines. Furthermore, defense expert Dr. Myslinski conceded that Dr. McClave's regression analysis is testable, generally accepted and reproducible. Regression and statistical analysis have been admitted in antitrust cases to prove injury and to determine damages." *Trauth Dairy*, 925 F. Supp. at 1252 (citations omitted).

below, however, this analysis of the regressions in *Conwood* is far too cursory.

a. Daubert's Four Factors

The problem with the district court's conclusion and reasoning is that the analyst did more than use linear regression to predict the value of a dependent variable. Regression is an apodictically valid tool for measuring how changes in one variable are associated with changes in other variables. The mathematics that generates a regression equation is sound, but whether a "regression rectangle" validly estimates damages involves additional considerations. The most important of these considerations is the underlying premise that resistance to anticompetitive conduct is linearly correlated with market share and is the explanation for the positive slope of the regression line. That theory is difficult to square with the kinds of factors listed in *Daubert*.³¹⁶ First, the resistance theory holds that the effects of anticompetitive conduct can be measured simply by identifying states in which market share grew less and attributing the entire difference to the challenged conduct. This theory appears to have been invented for use in the *Conwood* case and has never been tested—in that case or any other. This fact counts heavily against admissibility.³¹⁷ Second, "the theory or technique has [not] been subjected to peer review and publication."³¹⁸ The economic literature is devoid of any discussion of the resistance theory and the "regression rectangle" as a means of detecting and measuring harms from anticompetitive conduct. Third, because the theory and method have yet to be tested, the risks and magnitude of the errors that it yields are unknown. Finally, as indicated by the lack of published or other critical discourse about the approach, general acceptance in the scientific community is lacking. The record in

³¹⁶ See *supra* Section II.D.2.a.

³¹⁷ As the Court of Appeals for the Ninth Circuit pointedly observed on remand in *Daubert*, "[o]ne very significant fact to be considered is whether the experts are proposing to testify about matters growing naturally and directly out of research they have conducted independent of the litigation, or whether they have developed their opinions expressly for purposes of testifying." *Daubert v. Merrell Dow Pharms., Inc.*, 43 F.3d 1311, 1317 (9th Cir. 1995). Plainly, Leftwich devised the susceptibility theory and arrived at his opinion about it "expressly for purposes of testifying."

³¹⁸ See *Daubert*, 509 U.S. at 593.

Conwood contains no evidence that economists, statisticians, accountants, or finance professors accept the resistance theory and “regression rectangle” estimates of damages. The procedure bears no resemblance to the commonly accepted “before-and-after” and “yardstick” approaches that use meaningful control groups to separate the effects of anticompetitive conduct from other factors.³¹⁹

In short, to argue that the regression study in *Conwood* satisfies *Daubert* simply because it uses least-squares regression is tantamount to claiming that Ptolemy’s theory that the sun revolves around the earth is valid and generally accepted because these movements can be described by geometry. A meaningful application of *Daubert* requires verification of all the major premises of the analytical method, not just those at the highest level of abstraction.³²⁰ It is therefore appropriate to observe that no other expert in any antitrust case has used “rectangular regression” to infer causation or to estimate damages.

A narrower but still severe methodological flaw in the *Conwood* regression analysis is the use of “adjusted” market shares.³²¹ Using shares that were already adjusted to reflect the effect of illegal conduct to deduce the effect of that same conduct is extremely puzzling. Indeed, the resulting numbers are so lacking in probative value as to be excludable under the balancing test of Federal Rule 403. Unadjusted shares produce much lower damage estimates—\$155–238 million³²² instead of \$313–488 million.³²³ It therefore seems that the prejudicial effect of adjustment in this case substantially outweighs whatever minimal probative value the adjustment could have.

Yet another disturbing feature of the analysis is the use of the *difference* of 8.1 percentage points between the predicted share of 28.1% in 1997 and the arbitrary 1990 starting share of 20%. The

³¹⁹ These procedures compare the outcomes in situations where the conduct of interest is present to those in which it is absent. See, e.g., *In re Indus. Silicon Antitrust Litig.*, No. 95–2104, 1998 WL 1031507, at *2–3 (W.D. Pa. Oct. 13, 1998) (concluding that a “before-and after” regression satisfies *Daubert* and *Bazemore v. Friday*, 478 U.S. 385 (1986)).

³²⁰ See *supra* Section II.D.2.

³²¹ See *supra* note 308.

³²² Leftwich Trial Transcript, *supra* note 268, at 42.

³²³ See *supra* Section II.E.1.

share growth also can be expressed as the ratio of 28.1 to 20, or 1.405. Since this is the growth factor in a putatively resistant state, why not assume that, but for the allegedly unlawful acts, the low-share, susceptible states would have grown by the same factor of 1.405? This multiplicative adjustment would raise an initial 1% share to only 1.4% instead of 8.1%.

The assumption of additive growth implicit in the regression-rectangle damages estimate is a general feature of the method rather than a case-specific fact. Consequently, it is appropriate to apply *Daubert* and to demand a showing that the assumption is valid. In *Conwood*, no theoretical or empirical reason was offered to explain why growth would be additive rather than multiplicative.³²⁴ Even if the resistance theory were better established, the

³²⁴ Whether to use absolute disparities in market shares or growth rates is not a question of statistical theory. It is a question of economics—of how sales grow over time. The ideal way to estimate losses due to illegal conduct would be with a model that includes those variables that determine sales and prices. With such a model, one would set those variables to the levels that capture the relevant economic conditions in the affected market (with the illegal conduct removed). Market share would not be the dependent variable, and the question of whether to use the difference in share points or the percentage growth in share points would not arise.

Even without a meaningful economic model, one can examine trends. Suppose there are annual data on sales before and during the period of illegal conduct. If the pre-violation data are extensive enough, we might be able to discern the relationship between sales in one year and sales in the next. For example, sales might follow a straight line as a function of time:

$$Q_t = At, \quad (1a)$$

where A is the slope of the line and t is the year (1, 2, 3, . . .). In this case, the sales in any given year t are just the constant A added to the sales of the preceding year:

$$Q_{t+1} = Q_t + A. \quad (1b)$$

We would estimate A from the trend in the pre-violation period, then add it to the sales immediately before the violation period to estimate what sales would have been but for the violation in the first year of the violation period.

Using relative change (the ratio rather than the difference) would be a mistake in this situation. The percentage change in any year t is $100(Q_t - Q_{t-1})/Q_t = 100A/At = 100/t$. In year $t=10$, for instance, sales grow by 10%. In year 20, the growth is only 5% of the previous year's sales. Yet, the annual growth but for the illegal conduct is A , which does not change.

On the other hand, the data on sales in the pre-violation period might establish a different trend. Suppose sales grow by a fixed fraction B every year in this period:

$$Q_{t+1} = Q_t + BQ_t = (B+1)Q_t. \quad (2a)$$

This is an exponential growth pattern:

$$Q_t = Q_1(1 + B)^t. \quad (2b)$$

additive method of adjustment has not been validated, and it could be quite unreliable.

b. Daubert's Fit and Joiner's Nexus

Reliance on mathematics or statistics is not enough to satisfy *Daubert*. If such reliance were sufficient, then the plaintiffs' experts in both *Daubert* and *Joiner* would have been allowed to testify without further ado, for the experts in both these cases relied on statistical studies or analyses. In *Conwood*, the use of regression to estimate damages can be dismissed because it does not fit the problem,³²⁵ but the "fit" analysis adds nothing to the analysis of the validity of "rectangular regression" as a method for estimating damages.³²⁶ It is merely another way to say that although regression is a valid procedure for looking at the association between variables and for predicting the value of a dependent variable, the interpretation of differences in market share growth as the result of "resistance" to illegal conduct has no logical or scientific basis.

In *Joiner*, the Court wrote that heightened scrutiny encompasses not only the abstract methodology, but also the use of that methodology to reach specific conclusions. As discussed in Section II.C.2, in examining an expert's opinion based on standard statistical methods in epidemiology, the Court held that the opinion failed to satisfy *Daubert* because it was "connected to existing data only by the *ipse dixit* of the expert."³²⁷ In the end, there was "simply too

In this situation, estimating changes in sales by the relative growth in the preceding year would be correct. The sales should grow every year by 100B%.

The lesson from these examples is that there is no general rule as to which measure of change to use. It depends on the mechanism that produces the changes. Leftwich did not have a known trend line to use. He did not try to discern the pre-violation pattern. Rather, he used the equivalent of four points: (1) 1990 shares and (2) 1997 shares in the supposedly resistant, high-share states, and (3) 1990 and (4) 1997 shares in the susceptible, low-share states. Without further information, there is no way to say whether the more accurate estimate would come from projecting the simple difference in percentage points (as Leftwich did) or relative (percentage) growth.

³²⁵ Cf. McFadden's Brief, *supra* note 294, at 9 ("[T]o estimate the economic damages resulting from those [unlawful] acts, it is necessary to identify and quantify the prices and quantities that the plaintiff would have experienced but for the defendant's actions, and compare these hypothetical benchmark prices and quantities with the actual or as is prices and quantities.").

³²⁶ This reflects the fact that "fit" is part of the definition of validity. See *supra* text accompanying note 153.

³²⁷ *Joiner*, 522 U.S. at 146.

great an analytical gap between the data and the opinion proffered.”³²⁸ The phrase “*ipse dixit* of the expert,” much like the “gatekeeping” metaphor of *Daubert* itself, has great rhetorical force, but little analytical precision. Here, it is easy to dismiss the “rectangular regression” as an *ipse dixit* that cannot bridge the “gap between the data and the opinion proffered,” but the justification for these characterizations lies entirely in the preceding analysis of the putative validity of this novel procedure for estimating damages.

3. “Internal” Criticisms of the Regression

The major criticisms of the regression study in *Conwood* are “external” to the study. The problems with the “resistance theory” and the “rectangular regression” are present whether the regressions are performed impeccably or erroneously. These problems undermine the major premise that “resistance” can be presumed to be the explanation for variations in market share growth. Such external criticisms clearly affect the validity of this regression-based method for establishing damages.³²⁹

³²⁸ *Id.*

³²⁹ This conclusion does not turn on the internal-external distinction, which can be slippery. The objection that the resistance theory is invalid can be reformulated as a complaint that the regression ignores important variables. The choice of variables in a regression equation and the selection of a functional form (linear or nonlinear) sounds like an internal criticism because it bears on the execution of the particular regression. See *supra* note 78. However, attempting to infer causation and to measure the impact of a variable when this variable is not included in the regression (either directly or in the form of some adequate proxy) is a fundamental methodological flaw. It is a procedure that cuts across cases and that lacks validity (or, equivalently, “fit”).

More typical disputes over the failure to include possibly confounding variables should affect admissibility under Federal Rule 403 rather than Federal Rule 702. In extreme cases, exclusion will be warranted. See, e.g., *People Who Care v. Rockford Bd. of Educ.*, 111 F.3d 528, 537–38 (7th Cir. 1997) (determining that a study of an achievement gap among students of different races was inadmissible under *Daubert*); cf. *Smith v. Va. Commonwealth Univ.*, 84 F.3d 672, 677 (4th Cir. 1996) (*en banc*) (holding that omission of major variables precludes summary judgment); D.H. Kaye, *Statistical Evidence: How to Avoid the ‘Diderot Effect’ of Getting Stumped*, *Inside Litig.*, Apr. 1988, at 21 (proposing a standard for determining when an omitted variable weakens a regression to the point that it does not create a *prima facie* case of discrimination).

Other criticisms are internal to the study, and the propriety of judging admissibility under *Daubert's* heightened scrutiny for reliability is more debatable. For instance, the data set contains "outliers"—states that unduly influence the regression results.³³⁰ The amicus brief hammers hard at this point:

The difference between a finding by Dr. Leftwich of several hundred million dollars of damages and a finding of no damages is the inclusion in his model of a single anomalous data point, the data for Washington, DC ("DC").

Any reasonable statistical analysis would identify the DC point as one that does not fit the model. . . .

The question of whether the DC data point should be given the same weight as other data points is not an academic quibble. A fundamental step in producing a sound econometric analysis is to look for aberrant data that is [sic] either erroneous, highly variant, or does not fit the specified model. Any number of diagnostics would have identified the DC data point as an outlier that should either have been excluded from Dr. Leftwich's regressions or given less weight than other data. This is not an issue over which reasonable economists would differ.

. . . Dr. Leftwich's failure is not a subtle statistical mistake. This kind of failure to examine the impact of such an outlier would not be acceptable in an undergraduate econometrics class, let alone professional work.³³¹

At first blush, it is not clear whether the failure to attend to obvious outliers should be seen as a methodological flaw subject to heightened scrutiny or a case-specific defect in the implementation of a valid methodology to be screened only under the Federal Rule 403 standard.³³² It might seem that because the impact of the failure

³³⁰ Because the least-squares regression line is the line that minimizes the sum of the squared vertical deviations from the line to the data points, a single point that is far out of line can be very influential. For discussions of outliers, see Kaye & Freedman, *supra* note 15, at 137–38; Rubinfeld, *supra* note 19, at 199.

³³¹ McFadden's Brief, *supra* note 294, at 15–19.

³³² Of course, the exercise in classification is unnecessary under *Kumho*. Branding an analysis as "not acceptable in an undergraduate econometrics class, let alone professional work" is a kiss of death under *Kumho's* "same level of intellectual rigor" test. *Kumho*, 526 U.S. at 152.

to use standard regression diagnostics is a recurring issue that is likely to be opaque to most jurors, it should fall into the former category. Denominating a test for outliers as a “fundamental step in producing a sound econometric analysis” tugs in this direction, but the impact of an outlier on a regression line depends on the particular data set, and there is no simple corrective procedure.³³³ While it would be reasonable to place the burden of testing for outliers on the analyst, in some situations it may be valid to rely on least-squares regression results even though there is an outlier. The factors considered in assessing theories, methods, or propositions that serve as trans-case premises—considerations like error rates, scrutiny in the scientific literature, general acceptance, and the like—are not directed to addressing whether an expert’s conclusion as to handling an outlier in a particular regression is correct. Moreover, it should be feasible to convey to a jury the fact that a result vanishes when a single point is dropped and to convince them that reliance on so unstable a regression is foolhardy. As a result, the better approach might be to classify the treatment of an outlier as a case-specific matter.

Of course, this classification does not require the admission of results that are unstable. If the unreliability of a particular analysis is “not an issue over which reasonable economists would differ,”³³⁴ the probative value of the regression will be minimal and the testimony will not be worth the time and effort that would be required to educate the jury as to its limitations. From a doctrinal standpoint, however, there is no need to develop and apply special standards, and the danger of excluding a particular statistical study as lacking sufficient “intellectual rigor,” even though it is within the range of reasonable debate by experts, counsels against too stringent a level of screening under the gloss placed on *Daubert* in *Joiner* and *Kumho*.³³⁵

CONCLUSION

The recent efforts of the Supreme Court to develop special standards for the admissibility of scientific expert testimony have given

³³³ Rubinfeld, *supra* note 19, at 199.

³³⁴ McFadden’s Brief, *supra* note 294, at 18.

³³⁵ See *supra* note 243 and accompanying text.

new urgency to the search for solutions to longstanding problems or puzzles created by such screening. The broad contours of the *Daubert* trilogy are appealing, but the internal structure needs further bracing. Maintaining analytical clarity in this area of the law is challenging, and statistical and econometric evidence can serve as a crucible for testing theories of how courts should screen scientific evidence. Given the obstacles to lay comprehension of complex statistical analyses, statistical and econometric studies surely qualify for heightened scrutiny under *Frye* or *Daubert*. Thus, the recent trend toward careful screening of such evidence prior to trial is encouraging, but not all aspects of studies based on data of interest only to litigants should be strained through these filters. Phrases like “gatekeeping” and “intellectual rigor” are well and good, but heightened scrutiny should be reserved for methodology. Imperfections in the execution of a particular study should not result in exclusion unless they reduce the probative value to the point where it is substantially outweighed by the dangers of prejudice, confusion, and time-consumption.

APPENDIX

Table A1
Conwood's Market Shares by State in 1990 and 1997

(Source: Plaintiff's Exhibit No. 327.26)

State	Percent of Market		Gain or Loss (Share points)
	1990	1997	
Maine	24.2	27.2	3
Michigan	24	32.1	8
New Hampshire	21.5	26.9	5.5
Oregon	18.9	24.3	5.4
Illinois	18	20.4	2.5
Massachusetts	17.6	24.7	7.1
Kentucky	17.6	19.9	2.4
Wisconsin	17.2	27.7	10.5
South Carolina	17.1	23	5.9
Washington	16.7	21.3	4.5
Indiana	16	21.1	5.1
Minnesota	15.4	28.1	12.6
Ohio	15.3	19.2	3.8
Missouri	15	14.1	-0.9
Delaware	14.8	18.8	4
Vermont	14	18	4
Virginia	12.4	14.5	2.4
New Jersey	12.3	18.9	6.6
New York	12.2	17	4.8
North Carolina	11.9	15.9	3.9
Maryland	10.4	14.8	4.4
Louisiana	10.3	8.7	-1.6
Georgia	10.3	12.4	2.1
Tennessee	9.9	10.4	0.5
Iowa	9.9	12.4	2.5
Connecticut	9.8	16.9	7.1
Kansas	9.7	8	-1.7

Nebraska	8.5	8	-0.5
Alabama	7.7	8.1	0.4
California	7.7	9.3	1.6
Arkansas	7.6	8.2	0.6
Mississippi	7.4	8.1	0.7
District of Columbia	7.2	17.6	10.3
Colorado	7	7.9	0.9
Florida	6.3	7.7	1.4
South Dakota	5.8	7	1.2
North Dakota	5.8	10.2	4.4
Pennsylvania	5.3	7.5	2.1
Idaho	5.3	8.6	3.4
Arizona	5.2	6.3	1
Montana	5.1	5.4	0.3
West Virginia	5	5.1	0.1
Oklahoma	4.9	5.1	0.3
Utah	4.6	5.3	0.7
Nevada	4.5	5.7	1.2
Texas	4.5	4.8	0.3
New Mexico	3	3	0
Wyoming	2.4	2	-0.3
Rhode Island	1.4	15.2	13.7

*Table A2
Conwood's 1990 and 1997 Adjusted Market Shares by State*

(Source Plaintiffs' Exhibit No. 327.26)

State	Percent of Market		Gain or Loss (Share points)
	1990	1997	
Maine	24.2	30.3	6.1
Michigan	24	36.2	12.1
New Hampshire	21.5	26.9	5.5
Oregon	18.9	25.7	6.8
Illinois	18	22.6	4.7
Massachusetts	17.6	32.6	15
Kentucky	17.6	20.4	2.9
Wisconsin	17.2	33	15.8
South Carolina	17.1	23.5	6.4
Washington	16.7	24.2	7.4
Indiana	16	23.8	7.8
Minnesota	15.4	32.1	16.6
Ohio	15.3	20.4	5
Missouri	15	15.2	0.2
Delaware	14.8	18.8	4
Vermont	14	25	11
Virginia	12.4	16.2	3.8
New Jersey	12.3	18.9	6.6
New York	12.2	20.5	8.3
North Carolina	11.9	16.4	4.4
Maryland	10.4	15.4	5
Louisiana	10.3	10.7	0.4
Georgia	10.3	13.6	3.3
Tennessee	9.9	10.5	0.6
Iowa	9.9	12.4	2.5
Connecticut	9.8	24.3	14.5
Kansas	9.7	11.3	1.6
Nebraska	8.5	9.8	1.3
Alabama	7.7	8.6	0.9
California	7.7	10.7	3

Arkansas	7.6	9.5	1.9
Mississippi	7.4	8.1	0.7
District of Columbia	7.2	17.6	10.3
Colorado	7	9.3	2.3
Florida	6.3	8.6	2.2
South Dakota	5.8	7	1.2
North Dakota	5.8	12	6.2
Pennsylvania	5.3	9.9	4.5
Idaho	5.3	9.6	4.4
Arizona	5.2	8	2.7
Montana	5.1	10.2	5.1
West Virginia	5	5.5	0.5
Oklahoma	4.9	8.2	3.4
Utah	4.6	8.7	4.1
Nevada	4.5	5.7	1.2
Texas	4.5	6.2	1.7
New Mexico	3	5.5	2.6
Wyoming	2.4	2	-0.3
Rhode Island	1.4	15.2	13.7