

Oklahoma Law Review

Volume 52 | Number 2

1-1-1999

Relevance, Reliability, and Validity of Scientific Evidence (reviewing *Judging Science: Scientific Knowledge and the Federal Courts*, by Kenneth R. Foster & Peter W. Huber)

Robert Timothy Reagan

Follow this and additional works at: <https://digitalcommons.law.ou.edu/olr>



Part of the [Law Commons](#)

Recommended Citation

Robert T. Reagan, *Relevance, Reliability, and Validity of Scientific Evidence* (reviewing *Judging Science: Scientific Knowledge and the Federal Courts*, by Kenneth R. Foster & Peter W. Huber), 52 OKLA. L. REV. 291 (1999),
<https://digitalcommons.law.ou.edu/olr/vol52/iss2/7>

This Note is brought to you for free and open access by University of Oklahoma College of Law Digital Commons. It has been accepted for inclusion in Oklahoma Law Review by an authorized editor of University of Oklahoma College of Law Digital Commons. For more information, please contact darinfox@ou.edu.

BOOK REVIEW

RELEVANCE, RELIABILITY, AND VALIDITY OF SCIENTIFIC EVIDENCE

JUDGING SCIENCE: SCIENTIFIC KNOWLEDGE AND THE FEDERAL COURTS, by
Kenneth R. Foster & Peter W. Huber. Cambridge, Massachusetts: MIT Press. 333
pp. 1997.

*Reviewed by Robert Timothy Reagan**

In 1993, the United States Supreme Court declared federal judges to be "gatekeepers" charged with meaningful scrutiny of proffered scientific evidence.¹ Although the decision in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*² increases the intellectual burden on lawyers and judges, if we can meet the burden, more just results will follow.

Kenneth Foster and Peter Huber have offered an intriguing guide to meeting the intellectual burden of assessing the admissibility of scientific evidence.³ The structure of their 333-page book is excellent. Each of the nine chapters, except for the first and last, examines an important lesson of *Daubert*, and begins with text from Justice Blackmun's majority opinion. In addition, the book is peppered with sidebar boxes containing reprinted text from other authorities in science and law. The *Daubert* text and the sidebars are useful, because they are the raw data analyzed by Foster and Huber's book. Their inclusion permits ready comparison of the analysis to the textual data. Unfortunately, certain misconceptions about the subject matter make the book more provocative for scholars than useful for practitioners or judges.⁴

* A.B. 1980, Stanford University (Human Biology, Psychology); Ph.D. 1986, Harvard University (Psychology); J.D. 1993, Hastings College of the Law, University of California. Please direct correspondence to Dr. Tim Reagan, Research Division, The Federal Judicial Center, Thurgood Marshall Federal Judiciary Building, One Columbus Circle N.E., Washington, DC 20002-8003, (202) 502-4097, treagan@fjc.gov.

The views expressed in this review are those of the author, and not necessarily those of The Federal Judicial Center.

1. See *Daubert v. Merrell Dow Pharms. Inc.*, 509 U.S. 579, 589 n.7, 597 (1993); see also *Kumho Tire Co. v. Carmichael*, No. 97-1709, slip op. at 19 (U.S. Mar. 23, 1999); Stephen Breyer, *The Interdependence of Science and Law*, 82 JUDICATURE 24, 26 (1998).

2. 509 U.S. 579 (1993).

3. KENNETH R. FOSTER & PETER W. HUBER, JUDGING SCIENCE: SCIENTIFIC KNOWLEDGE AND THE FEDERAL COURTS (1997).

4. On a scale from one to four stars, the editors of *Jurimetrics* gave Foster and Huber's book four

Scientific Knowledge

Chapter 1 sets the stage for the *Daubert* analysis. It poses the overarching question: "[w]hat is scientific knowledge, and when is it reliable?"⁵ The title, "Scientific Knowledge," is taken from the text of Federal Rule of Evidence 702.

Rule 702. 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.⁶

Indeed, the *Daubert* Court declared Rule 702 to be the "specific Rule that speaks to the contested issue."⁷

Foster and Huber identify three federal rules "that bear directly upon scientific evidence in court"⁸ In addition to Rule 702, they identify Rule 403, which is used to exclude misleading evidence, and Rule 703, which permits expert witnesses to base their conclusions on hearsay. Curiously excluded are Rules 401 and 402,⁹ the fundamental rules of relevance.¹⁰

Rule 401. "Relevant evidence" means evidence having any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence.¹¹

Rule 402. All relevant evidence is admissible, except as otherwise provided by the Constitution of the United States, by Act of Congress, by these rules, or by other rules prescribed by the Supreme Court pursuant to statutory authority. Evidence which is not relevant is not admissible.¹²

All rules of evidence are derived from three basic rules. The first is the rule of relevance. To be admissible, evidence must be relevant to a material issue. The

stars. See Staff, Book Review, 37 JURIMETRICS J. 507, 507-08 (1997). I would give it two-and-a-half or three.

5. FOSTER & HUBER, *supra* note 3, at 1.

6. FED. R. EVID. 702 (emphasis added).

7. *Daubert*, 579 U.S. at 588.

8. FOSTER & HUBER, *supra* note 3, at 11.

9. See Scott Brewer, *Scientific Expert Testimony and Intellectual Due Process*, 107 YALE L.J. 1535, 1545-46 (1998) (including Rules 401 and 402).

10. See *Kumho Tire Co. v. Carmichael*, No. 97-1709, slip op. at 19 (U.S. Mar. 23, 1999) ("[S]cientific expert testimony . . . is admissible only if it is both relevant and reliable."); Paul C. Giannelli, *Daubert: Interpreting the Federal Rules of Evidence*, 15 CARDOZO L. REV. 1999, 2013-14 (1994).

11. FED. R. EVID. 401.

12. FED. R. EVID. 402.

second rule is the rule of prejudice. Relevant evidence should be excluded if it is likely to be misleading. The third rule is the rule of privilege. Relevant evidence is excluded if exclusion serves a policy goal more important than an accurate factual determination in a specific case.

The admissibility of scientific evidence is fundamentally a question of relevance.¹³ In *Daubert*, Justice Blackmun quoted from both Rules 401 and 402 to establish the "baseline" of the Court's analysis.¹⁴

Rule 401 defines relevance in terms of conditional probabilities. Evidence is relevant if the probability of a fact is different with the evidence from what it would be without the evidence. That means that the posterior probability (P) of a fact (F), given the evidence (E), is not equal to the prior probability of the fact. Thus, if:

$$P(F|E) \neq P(F)$$

then the evidence is relevant.

In *Daubert*, the proffered fact was a matter of general causation — that the drug Bendectin causes birth defects.¹⁵ The legal question was whether plaintiffs' scientific evidence made the posterior probability ($P(F|E)$) high enough to justify a jury verdict in plaintiffs' favor. The District Court and Court of Appeals determined that it did not.¹⁶

Of course, the lower courts relied on the older "general acceptance" standard set forth in *Frye v. United States*.¹⁷ In *Frye*, the Court of Appeals for the District of Columbia determined that a "systolic blood pressure deception test [had] not yet gained such standing and scientific recognition among physiological and psychological authorities as would justify the courts in admitting expert testimony deduced from the discovery, development, and experiments thus far made."¹⁸ Because the court had no personal experience with the accuracy of the test, and because its accuracy had not yet been established by the scientific community, the court had no way of determining whether the test was accurate. Therefore, the court could not determine whether the probability of truthfulness was related to the result of the test.¹⁹ Stated another way, the court could not determine whether the posterior probability ($P(F|E)$) was different from the prior probability ($P(F)$), and therefore whether the test result was relevant.

13. See Giannelli, *supra* note 10, at 2009-10; Shelley Storer, Note, *The Weight Versus Admissibility Dilemma: Daubert's Applicability to a Method or Procedure in a Particular Case*, 1998 U. ILL. L. REV. 231, 247.

14. See *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 587 (1993).

15. See *id.* at 582 ("[R]espondent moved for summary judgment, contending that Bendectin does not cause birth defects in humans.").

16. See *id.* at 583-84.

17. See *Frye v. United States*, 293 F. 1013, 1014 (D.C. Cir. 1923).

18. *Id.*

19. See Lee Loevinger, *Science as Evidence*, 35 JURIMETRICS J. 153, 157 (1995) ("Quite sensibly, the court concluded that it must determine the reliability and validity of the technology employed; and it decided that it should look to general acceptance by 'physiological and psychological authorities' as an indication of, or proxy for, reliability and validity.").

Because the admissibility of scientific evidence is fundamentally a question of relevance, it is a mistake to list only Rule 403 of the Federal Rules of Evidence as applicable. The basic question of scientific evidence is a Rule 401 question.²⁰

Fit

Rule 702 permits an expert witness to present scientific evidence if the evidence "will assist the trier of fact to understand the evidence or to determine a fact in issue"²¹ As Justice Blackmun observed, "[t]his condition goes primarily to relevance."²² Foster and Huber devote chapter 2 of their book to the component of relevance that Justice Blackmun called "fit," a term first used in this context by Judge Edward R. Becker of the Court of Appeals for the Third Circuit.²³

It was "fit" that got the *Daubert* plaintiffs thrown out of court on remand.²⁴ When the case first came before the Ninth Circuit, the question was one of *general* causation — does Bendectin cause birth defects generally?²⁵ On remand, the Ninth Circuit concluded that the plaintiffs might be able to produce enough evidence to survive summary judgment on the question of general causation,²⁶ but the plaintiffs could not produce enough evidence to survive summary judgment on the question of *specific* causation, that is, whether Bendectin caused the plaintiffs' individual birth defects.²⁷ The plaintiffs' case was entirely circumstantial.²⁸ The only proof they had that Bendectin caused their own birth defects was proof that their mothers took Bendectin during pregnancy and scientific evidence that Bendectin increases the risk of birth defects. In order to prove by a preponderance of the evidence that Bendectin caused *their* birth defects, the plaintiffs had to show that Bendectin at least doubled the risk of defects.²⁹ Otherwise, most occurrences of birth defects among babies whose mothers took Bendectin during pregnancy would be attributable to the underlying risk rather than to Bendectin. The plaintiffs had no scientific evidence that Bendectin had that great an effect on the risk of birth defects.³⁰

20. See, e.g., *United States v. Downing*, 753 F.2d 1224, 1234-35 (3d Cir. 1985).

21. FED. R. EVID. 702; see also *Daubert*, 509 U.S. at 591.

22. *Daubert*, 509 U.S. at 591.

23. See *id.* (citing *Downing*, 753 F.2d at 1242); see also Clifton T. Hutchinson & Danny S. Ashby, *Daubert v. Merrell Dow Pharmaceuticals, Inc.: Redefining the Bases for Admissibility of Expert Scientific Testimony*, 15 CARDOZO L. REV. 1875, 1912 (1994).

24. See *Daubert v. Merrell Dow Pharms., Inc.*, 43 F.3d 1311, 1320-22 (9th Cir. 1995) (*Daubert II*).

25. See *Daubert v. Merrell Dow Pharms., Inc.*, 951 F.2d 1128, 1130 (9th Cir. 1991) (questioning "[w]hether Bendectin is responsible for limb reduction defects") (*Daubert I*).

26. See *Daubert II*, 43 F.3d at 1320.

27. See *id.* at 1322; see also Michael D. Green, *Science is to Law as the Burden of Proof is to Significance Testing*, 37 JURIMETRICS. J. 205, 205 n.1 (1997) (distinguishing general causation and specific causation).

28. See *Daubert II*, 43 F.3d at 1320.

29. See *id.* at 1320-21; see also Green, *supra* note 27, at 221 (explaining why doubling of risk is required); Margaret A. Fergar, *Evidentiary Framework*, in FEDERAL JUDICIAL CENTER, REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 37, 95-96 (1994) (same).

30. See *Daubert II*, 43 F.3d at 1320-21 & nn. 12, 14.

The Ninth Circuit's analysis, however, is subject to possible misinterpretation, because there was nothing wrong with the fit between the proffered evidence and the material fact. The court explicitly agreed that proof of *general* causation can be circumstantial evidence of *specific* causation.³¹ The evidence was simply insufficient to meet the burden of proof.³²

Justice Blackmun's discussion of fit offered as an example astronomical evidence of whether the moon was full on a particular night.³³ Even if the evidence were quite accurate as to the fullness of the moon, the evidence would be admissible only to prove such facts as whether the night was dark, but not to prove such facts as whether people under the moon were especially prone to lunacy.³⁴

The question of fit is whether evidence proves what it purports to prove. Is evidence that the moon was full probative of whether the night was dark? Yes. Is it probative of whether people were crazy? No. Is evidence that Bendectin causes birth defects generally probative of whether it caused birth defects in a specific case? Yes. Was the evidence strong enough to survive summary judgment in *Daubert*? No.

In science, the question of whether evidence proves what it purports to prove is a question of "validity."³⁵ For example, does an IQ test really measure intelligence?³⁶ Do studies of thermal inertia on the surface of Mars tell us how rocky the surface is?³⁷ Scientists use the term "reliability" to discuss whether scientific evidence is *consistent*.³⁸ Do different IQ tests result in similar scores? Are scores similar regardless of who administers the test or when? Is Martian thermal inertia similar when measured from different spacecraft?

In science, reliability is weak circumstantial evidence of validity. A test cannot be a very good measure of an underlying phenomenon if the results are wildly inconsistent. Poor reliability, therefore, tends to negate validity. On the other hand, reliable results — meaning consistent results — are a good indication that the test is measuring *something*. Whether or not that something is the something of interest, however, is another question.

In law, evidence is considered "reliable" if it proves what it purports to prove.³⁹ That is why Justice Blackmun observed: "[i]n a case involving scientific evidence, *evidentiary reliability* will be based upon *scientific validity*."⁴⁰ This, really, is what

31. See *id.* at 1320 ("Modern tort law permits such proof . . .").

32. See *id.* at 1320-22.

33. See *Daubert I*, 509 U.S. 579, 591 (1993).

34. See *id.*

35. See *id.* at 590 n.9; Harold M. Ginzburg, *Use and Misuse of Epidemiologic Data in the Courtroom: Defining the Limits of Inferential and Particularistic Evidence in Mass Tort Litigation*, 12 AM. J.L. & MED. 423, 426 n.13 (1986).

36. See generally Ulrich Neisser et al., *Intelligence: Knowns and Unknowns*, 51 AM. PSYCHOLOGIST 77 (1996).

37. See generally Bruce H. Betts & Bruce C. Murray, *Thermally Distinct Ejecta Blankets From Martian Craters*, 98 J. GEOPHYSICAL RES. — PLANETS 11,043 (1993).

38. See Ginzburg, *supra* note 35, at 426 n.14.

39. See, e.g., Christopher Slobogin, *Psychiatric Evidence in Criminal Trials: To Junk or Not to Junk*, 40 WM. & MARY L. REV. 1, 33 (1998).

40. *Id.* at 33-34 (quoting *Daubert I*, 509 U.S. at 590 n.9); see also Joseph Sanders, *Scientific*

fit is all about. As Foster and Huber correctly observe: "[i]n law and science alike, fit is a matter of relevance — the extent to which an observation can be related, by a credible theory, to the issue at hand."⁴¹ Fit is viewed correctly not as a prong of the *Daubert* standard, but as another name for it.⁴²

The Court did "not presume to set out a definitive checklist or test. But some general observations are appropriate."⁴³ The four considerations the Court identified as important are: (1) falsifiability ("whether [the theory or technique] can be (and has been) tested");⁴⁴ (2) peer review ("whether the theory or technique has been subjected to peer review and publication");⁴⁵ (3) precision ("the known or potential rate of error");⁴⁶ and (4) general acceptance (which "can yet have a bearing on the inquiry").⁴⁷ This list of considerations fairly describes what scientists would consider in judging the probative value of scientific evidence.⁴⁸

Falsifiability

In a citation, the *Daubert* Court quoted philosopher of science Karl Popper, stating that "the criterion of the scientific status of a theory is its falsifiability, or refutability, or testability."⁴⁹ Foster and Huber, however, title their third chapter "Testability and Falsification."⁵⁰ A theory is falsifiable if it is *subject* to falsification, but it does not actually have to be falsified. If the theory is falsified, it is false. If it is not falsified, but could be by conceivable evidence, then the theory has merit.

Validity, Admissibility, and Mass Torts After Daubert, 78 MINN. L. REV. 1387, 1391 (1994).

41. FOSTER & HUBER, *supra* note 3, at 23; see also 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 Scientific Evidence*, 15 CARDOZO L. REV. 1799, 1801 (1994) ("Testimony that is not valid is not relevant."); David L. Faigman, *Mapping the Labyrinth of Scientific Evidence*, 46 HASTINGS L.J. 555, 568 (1995) ("[J]uries should only hear relevant evidence and scientifically invalid findings are irrelevant."); Hutchinson & Ashby, *supra* note 23, at 1912 ("This consideration relates primarily to relevance.").

42. Cf. Thomas D. Lyon & Jonathan J. Koehler, *The Relevance Ratio: Evaluating the Probative Value of Expert Testimony in Child Sexual Abuse Cases*, 82 CORNELL L. REV. 43, 45 n.14 (1996) ("Although *Daubert* discusses relevance it does so in terms of 'fit,' which is analogous to materiality and concerns the extent to which the expert testimony is 'tied to the facts of the case.'") (quoting *Daubert I*, 509 U.S. at 591).

43. *Daubert I*, 509 U.S. at 593; see also *Kumho Tire Co. v. Carmichael*, No. 97-1709, slip op. at (U.S. Mar. 23, 1999) ("*Daubert's* list of specific factors neither necessarily nor exclusively applies to all experts or in every case.>").

44. *Daubert I*, 509 U.S. at 593.

45. *Id.*

46. *Id.* at 594.

47. *Id.*

48. See Hutchinson & Ashby, *supra* note 23, at 1886-87. See *id.* at 1909-12 for a discussion of other relevant factors.

49. *Daubert I*, 509 U.S. at 593; see also FOSTER & HUBER, *supra* note 3, at 36 (quoting *Daubert I*, 509 U.S. at 593); *id.* at 40 (quoting Karl Popper).

50. FOSTER & HUBER, *supra* note 3, at 37 (emphasis added).

Foster and Huber's confusion between falsifiability and falsification is likely to confuse readers who do not come to their text with an understanding of the issue. The confusion manifests itself in various ways. Foster and Huber observe: "[a] theory, even though it may be scientific in its basic thrust, is not a very good theory at all if it is so loosely phrased that it cannot be proved wrong — if it is in fact wrong."⁵¹ This statement is true enough up to the dash, but the last clause is meaningless. Only a falsifiable theory can be right (true) or wrong (false). A theory too loosely phrased to be tested can be neither right nor wrong in a scientific sense. It "isn't even good enough to be wrong."⁵²

Foster and Huber assert that "[m]ost scientists, most of the time, still labor to confirm, not to falsify, previously articulated theories."⁵³ This statement reveals confusion because it suggests that the falsifiability principle means that scientists should be engaged in the destructive endeavor of proving each other wrong.⁵⁴ It is true that scientists often prove hypotheses by disproving rival hypotheses, but the rival hypotheses are often straw men or "null hypotheses."⁵⁵

Foster and Huber conclude their third chapter with the beginning of a recurring attack on the individual experts who appeared in the *Daubert* case.⁵⁶ The attack is largely unfair. Dr. Shanna Swan is criticized for testifying that epidemiological studies have not proven that Bendectin does not cause birth defects. Foster and Huber admit that the testimony is correct, but fault Dr. Swan for making the statement because it might mislead a layperson.⁵⁷ The admissibility of evidence is not a matter to be determined by witnesses, and expert witnesses should not be vilified because the party for whom they testify has lost the case.

Precision

Foster and Huber's chapter on precision ("Errors in Science") provides a comprehensive discussion of statistical significance testing. They provide very

51. *Id.* at 39.

52. Breyer, *supra* note 1, at 26.

53. FOSTER & HUBER, *supra* note 3, at 48.

54. For example, physicist and Nobel-laureate Steven Weinberg has observed that "one can imagine a category of experiments that *refute* well-accepted theories, theories that have become part of the standard consensus of physics. *Under this category I can find no examples whatever in the past one hundred years.*" STEVEN WEINBERG, DREAMS OF A FINAL THEORY: THE SCIENTIST'S SEARCH FOR THE ULTIMATE LAWS OF NATURE 129-30 (1993). Professor Weinberg noted that falsifying data usually results in theory modification first and only results in a shift to another theory after the modifications become too cumbersome as a matter of taste.

[A] scientific theory can never be absolutely ruled out by experimental data because there is always some way of manipulating the theory or the auxiliary assumptions to create an agreement between theory and experiment. At some point one simply has to decide whether the elaborations that are needed to avoid conflict with experiment are just too ugly to believe.

Id. at 125.

55. FOSTER & HUBER, *supra* note 3, at 49.

56. *See id.* at 63-68.

57. *See id.* at 67.

useful information on significance levels, confidence intervals, type I and type II error, and meta-analysis. The chapter includes a very instructive sidebar that shows how even measurements whose precision we now take for granted were once known with considerably less precision.⁵⁸ The speed of light is used as an example. This example also shows how experts can underestimate their own lack of precision or rate of error.⁵⁹

This chapter concludes with a gratuitous attack on the *Daubert* plaintiffs' expert, Dr. Alan Done.⁶⁰ Dr. Done's testimony was criticized because: "Shanna Swan was unable to confirm or even explain numerous aspects of Done's analysis."⁶¹ Dr. Done's testimony may or may not have been flawed. Another expert's inability to confirm or explain it is not dispositive of the issue.

Reliability and Validity

One of the most significant intellectual contributions of Justice Blackmun's *Daubert* opinion was his instruction that: "[i]n a case involving scientific evidence, *evidentiary reliability* will be based upon *scientific validity*."⁶² If evidentiary reliability and scientific validity are equivalent, then Foster and Huber's decision to devote separate chapters to the two concepts — Chapter 5 on "Reliability" and Chapter 6 on "Scientific Validity" — seems a poor choice.

The terms "reliability" and "validity" are terms of art in science. That means that looking to *The Oxford English Dictionary* for general dictionary definitions, as Foster and Huber do, is unlikely to achieve a correct result, and does not in this case.⁶³ Behavioral scientists, Dr. Robert Rosenthal and Dr. Ralph Rosnow, define "reliability" as: "[t]he degree to which observations are consistent or stable."⁶⁴ They define "validity" as: "[t]he degree to which we observe what we purport to observe."⁶⁵ The degree to which scientific evidence proves what it purports to prove is a matter of scientific validity, but in court it is a matter of evidentiary reliability.⁶⁶

Foster and Huber include an interesting discussion of Bayes' theorem. According to Bayes' theorem, the following is true:⁶⁷

$$\frac{P(F|E)}{P(\sim F|E)} = \frac{P(F)}{P(\sim F)} \cdot \frac{P(E|F)}{P(E|\sim F)}$$

58. *See id.* at 94-95.

59. *See id.* at 90.

60. *See id.* at 104-07.

61. *Id.* at 104.

62. *Daubert I*, 509 U.S. 579, 590 n.9 (1993).

63. *See* FOSTER & HUBER, *supra* note 3, at 111, 138.

64. ROBERT ROSENTHAL & RALPH L. ROSNOW, *ESSENTIALS OF BEHAVIORAL RESEARCH* 479 (1984) (glossary).

65. *Id.* at 482.

66. *See Daubert I*, 509 U.S. at 590 n.9.

67. *See, e.g.*, MICHAEL O. FINKELSTEIN & BRUCE LEVIN, *STATISTICS FOR LAWYERS* 93 (1990).

As in the relevance formula at the beginning of this review, "P" stands for the probability of, "F" stands for the fact at issue, and "E" stands for the evidence. The tilde "~" means "not," so "~F" means "not F," or that "F," the fact, is not true.

Bayes' theorem shows that the posterior probability of a fact (the probability of the fact taking into account the evidence) will be different from the prior probability of a fact (the probability of the fact without knowing the evidence) when the evidence is more or less likely depending upon whether the fact is true or false. In symbols, $P(F|E)$ is different from $P(F)$ when $P(E|F)$ is different from $P(E|\sim F)$.⁶⁸ This means that evidence is relevant if the likelihood ratio for the evidence conditioned on the fact is different from one.⁶⁹

Foster and Huber's discussion of Bayes' theorem is important because it refutes the fallacy of the transposed conditional — the false assumption that $P(F|E)$ equals $P(E|F)$.⁷⁰ Foster and Huber correctly show that a diagnostic test with a 98% accuracy rate can nevertheless result in a positive result that is right only 15% of the time if the prior probability of the characteristic tested is low enough.⁷¹ But Foster and Huber commit a common error in describing this phenomenon as a low "predictive value of the test."⁷² This is merely low predictive value of a *positive result*. The vast majority of negative results will be accurate.

Foster and Huber also err when they refer to prior probabilities as "initial guesses."⁷³ Prior probabilities are simply the base rate probabilities without taking into account the evidence on which the posterior probabilities are based.⁷⁴ It is a mistake to overinterpret the time metaphor used in the terms. It is simply not true that "[a]s more evidence comes in, each scientist changes his or her personal probabilities for each hypothesis and assigns new values, which are called *posterior probabilities*."⁷⁵ Epidemiologists, for example, do not compute the probability that Bendectin causes birth defects. They compute the probability that evidence such as that observed would result if Bendectin *did not* cause birth defects.⁷⁶

68. To see how this is so, keep in mind that $P(\sim F|E) = 1 - P(F|E)$ and $P(\sim F) = 1 - P(F)$.

69. See BERNARD ROBERTSON & G.A. VIGNAUX, *INTERPRETING EVIDENCE: EVALUATING FORENSIC SCIENCE IN THE COURTROOM* 21 (1995); Stephen E. Fienberg & Mark J. Schervish, *The Relevance of Bayesian Inference for the Presentation of Statistical Evidence and for Legal Decisionmaking*, 66 B.U. L. REV. 771, 773-76 (1986); Lyon & Koehler, *supra* note 42, at 46-47. Professors Lyon and Koehler describe this likelihood ratio as a "relevance ratio." *Id.*

70. See, e.g., Richard D. Friedman, *Assessing Evidence*, 94 MICH. L. REV. 1810, 1837 & n.64 (1996) (discussing the fallacy); Robert Timothy Reagan, *Supreme Court Decisions and Probability Theory: Getting the Analysis Right*, 76 U. DET. MERCY L. REV. (forthcoming 1999) (same).

71. See FOSTER & HUBER, *supra* note 3, at 116.

72. *Id.* (emphasis added).

73. *Id.* at 122.

74. See, e.g., FREDERICK MOSTELLER & DAVID L. WALLACE, *APPLIED BAYESIAN AND CLASSICAL INFERENCE: THE CASE OF THE FEDERALIST PAPERS* 49 (2d ed. 1984).

75. FOSTER & HUBER, *supra* note 3, at 122.

76. See, e.g., Franco Taroni & Colin Aitken, *Forensic Science at Trial*, 37 JURIMETRICS J. 327, 329 (1997). Actually, epidemiologists do not really compute *causation* probabilities at all. They compute *correlation* probabilities. Their method is observational, not experimental. See, e.g., HANS ZEISEL & KAVID KAYE, *PROVE IT WITH FIGURES: EMPIRICAL METHODS IN LAW AND LITIGATION* 45 (1997).

Peer Review

Foster and Huber's very interesting chapter on "Peer Review and the Scientific Community" pierces many of the myths of peer review. The courts' frequent requirement that an expert's testimony be previously published in a peer-reviewed academic journal is quite a puzzle. Most legal scholarship relied upon by the courts is not published in peer-reviewed journals, but in student-edited journals.⁷⁷

The peer review in the selection of an article for publication does not imply that the scientific community stands behind the article's findings.⁷⁸ In an ideal case, it implies only that three to five scientists judged the method to be competent and the analysis to be interesting.⁷⁹

Prejudice

Although the admissibility of scientific evidence is fundamentally a question of relevance, the question of prejudice will often be a factor to consider. Foster and Huber wisely include a chapter on "Prejudicing, Confusing, or Misleading the Jury." As they remind us, "[t]his problem [prejudice] is entirely distinct from the ones discussed elsewhere" in their book.⁸⁰

Foster and Huber's discussion of sophistry and scientism illuminates many of the issues that must be considered in determining whether scientific evidence, although sufficiently valid to be relevant, is, nevertheless, so likely to be overinterpreted as to be unfairly prejudicial. We do not yet have an easy test to make this determination, but Foster and Huber offer useful considerations.

Conclusion

Foster and Huber conclude their book with a thirty-three-page summary chapter which is, in part, merely a shorter version of the book itself. However, it also includes five new sidebar quotations. As a matter of organization, this chapter is too short and includes too much new material.

Peter Huber discusses "Bendectin in the Press: The Misreporting of Law and Science" in an article attached to the book as Appendix A. This addendum is largely a complaint that the press treated Merrell Dow badly during the course of the Bendectin litigation. Some observations, such as that Linda Greenhouse of the *New York Times* is one of the best legal reporters working,⁸¹ are true enough. But the prediction made by the San Diego *Union-Tribune* that the Supreme Court's decision in *Daubert* meant that the case would go to trial is not a good example of

77. See, e.g., Lars Noah, *Sanctifying Scientific Peer Review: Publication as a Proxy for Regulatory Decisionmaking*, 59 U. PITT. L. REV. 677, 694 (1998).

78. See, e.g., Randolph N. Jonakait, *The Meaning of Daubert and What That Means For Forensic Science*, 15 CARDOZO L. REV. 2103, 2111 (1994).

79. See generally Hutchinson & Ashby, *supra* note 23, at 1900-05; Noah, *supra* note 77, at 678.

80. FOSTER & HUBER, *supra* note 3, at 209.

81. See *id.* at 263-64.

poor journalism.⁸² Indeed, a jury trial was a possibility after the Supreme Court's decision. The Ninth Circuit's Court of Appeals, however, threw out the case on a ground not considered previously in the litigation.⁸³

Foster and Huber's analysis of *Daubert* and its implications is provocative and interesting. Although it will inspire useful thinking by sophisticated readers, it is likely to mislead the unsophisticated on many issues. Therefore, it should not be relied upon heavily in the course of litigation.

82. *See id.* at 264.

83. *See Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 43 F.3d 1311, 1320-22 (9th Cir. 1995) (affirming Merrell Dow's summary judgement on the ground that plaintiffs could not prove that Bendectin at least doubted the risk of birth defects).

