Essay

Finding the Error in *Daubert*

MARK HAUG* AND EMILY BAIRD**

The Supreme Court in Daubert v. Merrell Dow Pharmaceuticals, Inc. *laid down the standard for admissibility of expert testimony. We believe the best standard is simpler than the one chosen by the Court: The* Daubert *standard really is about discerning the trustworthiness of expert, and trustworthiness is best determined through an expert's accounting of the error within his testimony. Lower courts have struggled with the* Daubert *standard. We offer evidence of the problem and propose a new standard that would capture the essence of* Daubert *but significantly simplify its application.*

* Ph.D., J.D., School of Business Teaching Fellow at the University of Kansas. We wish to thank Lauren Moreno for her suggestions and the editorial staff at *Hastings Law Journal* for their time and assistance in editing our paper.

** Masters in Professional Accounting Candidate, The University of Texas at Austin.

Table of Contents

Introduction

In federal courts, expert evidence enjoys two freedoms other evidence does not: permissible hearsay and opinion testimony.[1] Accordingly, the rule for admissibility of expert testimony[2] has a significant impact on litigation. Many high-stakes cases turn on "the battle of the experts," engaging counsel and judges alike to apply the *Daubert* standard. We believe this standard is difficult to apply. Our Essay proposes a new standard that simply examines the three types of error that exist in the scientific endeavor. Such a bold proposal as ours, of course, requires a credible argument as to why current thinking is flawed and examining the past and present struggles with admitting expert opinion.

Our proposal is consistent with *Daubert*'s[3] motivation and Federal Rule of Evidence 702.[4] In our belief, it is also more faithful to Federal Rule of Evidence 403,[5] the Fifth,[6] Seventh,[7] and Ninth Amendments,[8] and

---

1. Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 592 (1993) ("Unlike an ordinary witness, an expert is permitted wide latitude to offer opinions, including those that are not based on firsthand knowledge or observation. Presumably, this relaxation of the usual requirement of firsthand knowledge—a rule which represents 'a most pervasive manifestation of the common law insistence upon 'the most reliable sources of information.'" (citations omitted) (quoting Fed. R. Evid. 602 advisory committee's note) (internal quotation marks omitted)).

2. *See generally* Fed. R. Evid. 702; Kumho Tire Co. v. Carmichael, 526 U.S. 137 (1999); Gen. Electric Co. v. Joiner, 522 U.S. 136 (1997); *Daubert*, 509 U.S. 579; Frye v. United States, 293 F. 1013 (D.C. Cir. 1923).

3. 509 U.S. 579.

4. Fed. R. Evid. 702.

5. Fed. R. Evid. 403.

6. U.S. Const. amend. V.

7. U.S. Const. amend. VII.

promotes expedient litigation. We propose the following rule:

> If an expert can account for the measurement error, the random error, and the systematic error in his evidence, then he ought to be permitted to testify. On the other hand, if he should fail to account for any one or more of these three types of error, then his testimony ought not be admitted.

Our rule simply asks an expert for an accounting of error in his evidence, which, of course, is the primary interest in adjudicating opinion testimony. If an expert cannot account for each of science's three errors,[9] then it is self-evident that the expert himself is uncertain as to the degree of credibility that ought to be placed upon his opinion. A jurist ought to bar the expert from testifying, as the expert is either not an able expert or his testimony is no longer grounded in expert judgment but has crossed into speculation. If on the other hand, the expert can account for all three errors, then his opinion contains an objective degree of credibility. The expert's opinion, accounting for all three kinds of error, ought to be the subject of "vigorous cross-examination,"[10] because his opinion would provide a common and scientifically acceptable basis and could therefore either be rebutted based on the facts of the case, or challenged as an expert opinion.

To explain our proposal, we begin with a short review of *Daubert* and its progeny.[11] In our review, we consider some of the ways courts have managed the four nonexclusive criteria[12] promulgated in *Daubert* for the purpose of assessing evidentiary reliability[13] and, ultimately, admissibility.[14] We consider data which provide a look at how these criteria have been implemented in the lower courts. Following this, we provide a taxonomy of error: an explanation and illustration of three types of error in science. In our conclusion, we show how our proposal satisfies all four criteria, is more expedient,[15] and is more manageable for the gatekeeper.[16]

---

8. U.S. Const. amend. IX.

9. *See infra* Part III; *see also* Mark Haug, *Minimizing Uncertainty in Scientific Evidence*, *in* Scientific Evidence Review: Monograph No.7, at 87, 88–89 (Cynthia H. Cwik & Helen E. Witt eds., 2006).

10. Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 596 (1993).

11. Kumho Tire Co. v. Carmichael, 526 U.S. 137 (1999); Gen. Electric Co. v. Joiner, 522 U.S. 136 (1997).

12. *Daubert*, 509 U.S. at 593–94.

13. *Id.* at 590.

14. *Id.* at 592.

15. "Scientific conclusions are subject to perpetual revision. Law, on the other hand, must resolve disputes finally and quickly." *Id.* at 597.

16. *See generally* Tal Golan, Laws of Men and Laws of Nature: The History of Scientific Expert Testimony in England and America (2004) (examining the past and present struggles with admitting expert opinion).

I. *DAUBERT*'S FOUR FACTORS TO ASSESS EVIDENTIARY RELIABILITY

The Court in *Daubert* set out to determine whether the Federal Rules of Evidence superseded the rule for admissibility of expert opinion established in *Frye v. United States*[17] seventy years earlier.[18] Daubert alleged that Merrell Dow's medication, Bendectin, caused birth defects.[19] Applying the *Frye* standard of general acceptance,[20] the lower courts had determined that it was not generally accepted that Bendectin caused birth defects.[21] Daubert argued that the Federal Rules of Evidence, however, suggested a more expansive view of admissibility of expert opinion than *Frye*.[22] The Court agreed.[23]

The Court in *Daubert* proposed the following criteria for determining the admissibility of expert opinion under the Federal Rules of Evidence: testability,[24] peer review,[25] known or potential rate of error,[26] and general acceptance.[27] Although much has been written concerning *Daubert*, we find no conclusive material on *Daubert*'s factor of "known or potential error rate."[28] Three possible reasons include: (1) it is difficult to define, but "we know it when we see it"; (2) it is merely a detail of "evidentiary reliability" and therefore, does not warrant such attention; or, (3) it is too difficult to implement.

Whatever the reasons may be, this factor is worthy of consideration, at the very least, to alert the unsuspecting jurist of the perils of error identification and quantification. Our purpose, however, is far bolder: We believe that within this factor lies the crux of admitting opinion testimony. Whether for the minimal purpose of illuminating a mere factor or for the grander purpose of a new rule of admissibility, our

17. 293 F. 1013, 1014 (D.C. Cir. 1923).

18. *Daubert*, 509 U.S. at 585–87.

19. *Id.* at 582.

20. "General Acceptance" refers to the methods, practices, and body of knowledge generally accepted within the scientific discipline under consideration. *See id.* at 584.

21. *Id.* at 583–85 (citing *Frye*, 293 F. at 1014).

22. *Id.* at 587–89.

23. *Id.* at 587.

24. *Id.* at 593 ("Ordinarily, a key question to be answered in determining whether a theory or technique is scientific knowledge that will assist the trier of fact will be whether it can be (and has been) tested.").

25. *Id.* at 593–94 ("Another pertinent consideration is whether the theory or technique has been subjected to peer review and publication. . . . [S]ubmission to the scrutiny of the scientific community is a component of 'good science,' in part because it increases the likelihood that substantive flaws in methodology will be detected.").

26. *Id.* at 594 ("Additionally, in the case of a particular scientific technique, the court ordinarily should consider the known or potential rate of error and the existence and maintenance of standards controlling the technique's operation." (citation omitted)).

27. *Id.* ("Finally, 'general acceptance' can yet have a bearing on the inquiry. . . . '[A] known technique which has been able to attract only minimal support within the community' may properly be viewed with skepticism." (quoting United States v. Downing, 753 F.2d 1224, 1238 (3d Cir. 1985))).

28. *Id.* at 580.

objective is to enable legal professionals to use sound judgment with respect to scientific error and thereby provide greater clarity and predictability to *Daubert* hearings. In pursuit of our objective, we endeavor to convince the reader that *Daubert*'s list of four nonexclusive factors,[29] as well as any other factor illuminating evidentiary reliability of scientific evidence, may be subsumed into an able analysis of error.

We begin with our belief that the United States Supreme Court did not know exactly what to do with the "known or potential rate of error" factor. For example, the Court provides imprecise support for its decision. In footnote 9, the Court supplies, in part, "the difference between accuracy, validity, and reliability may be such that each is distinct from the other by no more than a hen's kick."[30] Indeed, the Court has technically mislabeled this factor.[31] Superficially, the factor is seemingly well designed, with a strong whiff of the "stuff" of science. The general acceptance factor, the peer review factor, and the testability factor are all qualitative judgments, while the "error" factor is seemingly an invitation for some quantitative—and therefore difficult—indicia. The "error" factor invites a number of variations laid upon the "gatekeeper," such as: Must the error rate be known with exactitude (rather than potential)?[32] Shall the error rate be held to some predetermined standard

---

29. *Id.* at 593–94.

30. *Id.* at 590 n.9 (quoting James E. Starrs, Frye v. United States *Restructured and Revitalized: A Proposal to Amend Federal Evidence Rule 702*, 26 JURIMETRICS J. 249, 256 (1986)) (internal quotation marks omitted). As we have set out to show in this Essay, the issues of "accuracy," "validity," and "reliability" are independent topics and are quite different in their diagnosis and prognosis.

31. The Court's persistence with the term "rate of error" is a hint that the Court is uncertain what to do with the idea of error: For all of the Court's careful and precise thought, its use of the term "rate" is either grossly misplaced or contemplates something that needs further description. The term "rate" refers to a ratio with some quantity per unit of time. *See, e.g.*, Regina C. Elandt-Johnson, *Definition of Rates*, 102 AM. J. EPIDEMIOLOGY, 267, 268 (1975). Scientific errors, however, are generally understood as proportions—a quantity where the numerator is contained in the denominator. *Id.*

32. *See, e.g.*, Sorto-Romero v. Delta Int'l Mach. Corp., No. 05-CV-5172 (SJF) (AKT), 2007 U.S. Dist. LEXIS 71588, at *22–23 (E.D.N.Y. Sept. 24, 2007) (finding that plaintiff's proferred engineering expert, who would testify that the wood-carving tool that caused injury was defective, was unreliable for reasons including that an error rate could not be produced); Rabozzi v. Bombardier, Inc., No. 5:03-CV-1397 (NAM/DEP), 2007 U.S. Dist. LEXIS 21724, at *7, *8, *20 (N.D.N.Y. Mar. 27, 2007) (finding a civil engineering expert's testimony regarding a jet boat design to be inadmissible, as none of the four *Daubert* criteria were met, including that the expert did not provide a known rate of error); Roane v. Greenwich Swim Comm., 330 F. Supp. 2d 306, 309, 319 (S.D.N.Y. 2004) (finding that the proffered expert in the field of mechanical engineering was precluded from testifying, in part due to the inability to produce a rate of error); Nook v. Long Island R.R. Co., 190 F. Supp. 2d 639, 641–42 (S.D.N.Y. 2002) (granting motion to exclude plaintiff-employee's expert testimony from an industrial hygienist on the grounds that the expert did not uphold the *Daubert* standards and could not offer a known rate of error); United States v. Towns, 19 F. Supp. 2d 67, 70–72 (W.D.N.Y. 1998) (allowing but limiting the testimony of a clinical psychology expert regarding defendant's mens rea and his claim of mental illness being the cause of his attempted bank robbery, because the proffer asserted that an error rate would be provided by said expert). Similarly, in *Phillips v. Raymond Corp.*, defendants moved to exclude the plaintiff's expert in the field of biomechanics. 364 F. Supp. 2d 730, 732–33 (N.D. Ill. 2005). The court granted the motion, finding that the expert had not reliably tested his assertions,

among scientists or jurists?[33] Is it sufficient that the expert can calculate the known or potential rate of error?[34] Is it sufficient that the gatekeeper can assess the known or potential rate of error?[35] The factor invites other

because the error rate was referenced as being unknown; in order to have provided an accurate error rate, the expert would have had to use "retrospective analysis," which he did not do. *Id.* at 740–41.

33. *See, e.g.*, United States v. Microtek Int'l Dev. Sys. Div., Inc., No. 99-298-KI, 2000 U.S. Dist. LEXIS 2771, at *2, *10–13, *15 (D. Or. Mar. 10, 2000) (excluding data from polygraph test because the error rate was referenced as being low in highly controlled studies, but when applied in the "real world," was generally much higher, and because data showed that, on average, false negatives occurred 5% of the time and false positives, 10% of the time); Buckman v. Bombardier Corp., 893 F. Supp. 547, 556–57 (E.D.N.C. 1995) (allowing the testimony of an expert offered to provide a comparative test studying the engine of a Sea Doo in choppy water, but limiting the testimony to that of a lay person, rather than an expert, because the test performed was found to be unreliable, partially due to a high estimated rate of error). Another example is *In re Ephedra Products Liability Litigation*, 393 F. Supp. 2d 181, 184 (S.D.N.Y. 2005), which consolidated numerous civil actions claiming that the consumption of dietary supplements containing Ephedra resulted in injury or death. The defendants moved to exclude expert testimony presented by the plaintiffs. *Id.* *P*-values and confidence intervals were identified as the two ways that epidemiological studies are statistically evaluated. *Id.* at 191. The experts' study offered that "the fivefold increased rate of hemorrhagic stroke among study participants who took more than 32 mg of ephedra on the day before the case's stroke has a '95% confidence interval of 0.84 to 41.33.'" *Id.* The statistical significance of the study was therefore found to be insufficient, as that "interval includes the value 1.0 (which would mean no increased risk)." *Id.* The referenced *p*-value was .07, which would assume a "less than a one-in-14 chance of [an association between ephedra and stroke] being due to a sampling error." *Id.* at 191 n.7.

34. *See, e.g.*, Albert v. Jordan, Nos. 05CV516, 05CV517, 05CV518, 05CV519, 2007 U.S. Dist. LEXIS 92025, at *2–3 (W.D. La. Dec. 14, 2007) (denying plaintiffs' motion to exclude the testimony of an expert in the field of vocational rehabilitation, while also finding that the expert provided reliable information, and noting that the rate of error was known without further addressing it); Benkwith v. Matrixx Initiatives, Inc., 467 F. Supp. 2d 1316, 1326, 1330, 1332 (M.D. Ala. 2006) (granting defendant's motion to exclude testimony of an expert in the field of epidemiology regarding a spray causing plaintiff's anosmia, because the opinions had not been tested and a rate of error could not be provided).

35. For example, in *Bone Shirt v. Hazeltine*, 461 F.3d 1011, 1020 (8th Cir. 2006), the Eight Circuit affirmed the district court's decision to allow testimony of a regression analyst commissioned by plaintiffs' Native American voters in order to show that the redistricting of the legislative area caused a bloc in the Native American vote. In a concurring opinion, however, Judge Gruender notes some discomfort with the testimony due to *Daubert*'s known or potential rate of error factor, explaining that it is "difficult to weigh this factor in *Daubert*'s analysis if 'the effect of that error is unknown.'" *Id.* at 1026 (Gruender, J., concurring) (quoting Bone Shirt v. Hazeltine, 336 F. Supp. 2d 976, 1002 (D.S.D. 2004)). For other examples, *see e.g.*, United States v. Leblanc, 45 F. App'x 393, 398, 400 (6th Cir. 2002) (finding evidence was inadmissible when plaintiff sought to introduce testimony of a child psychologist regarding the susceptibility of children to coercive interrogation in a case of sexual assault, because the opinion relied on "'soft science' . . . in which 'error is . . . rampant'" (quoting the district court)); United States v. Sullivan, 246 F. Supp. 2d 696, 698–99 (E.D. Ky. 2003) (admitting expert's testimony on the unreliability of eyewitness testimony in a *Daubert* hearing, and finding that the factor of error rate was not relevant as the study in question relied on the accuracy of observations made by other people based on the order in which photographs were placed); Ambrosini v. Upjohn Co., No. 84-3483 (NHJ), 1995 U.S. Dist. LEXIS 21318, at *16, *22–24 (D.D.C. Oct. 18, 1995) (finding that plaintiff's teratology expert was not permitted to testify, because the methodology used was found to be unreliable and could not yield an accurate error rate). Similarly, in *Miller v. Pfizer, Inc.*, where plaintiffs filed suit against a drug manufacturer after their son committed suicide shortly after taking a drug prescribed to treat depression, the court affirmed the district court's ruling rejecting plaintiff's expert—whose opinion would have provided evidence correlating the suicide and drug consumption.

combinations of these questions as well.[36] The Court's language simply provides:

> Additionally, in the case of a particular scientific technique, the court ordinarily should consider the known or potential rate of error, and the existence and maintenance of standards controlling the technique's operation.[37]

With such minimal guidance from the Supreme Court, the lower court on remand found two of the four factors—one being the rate of error—"difficult or impossible to apply to the expert testimony in th[e] case."[38]

More recently, Federal Rule of Evidence 702 was amended, inter alia, to reflect the Court's reasoning in *Daubert* and its progeny, including *Joiner* and *Kumho Tire*:

> 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

---

356 F.3d 1326, 1330, 1334 (10th Cir. 2004). Concerning the rate of error, the court held that the test lacked sufficient sample size, controls, and interaction with study participants to be based on sound facts. *Id.* As a final example, the plaintiff in *Quiet Tech. DC-8, Inc. v. Hurel-Duboi U.K., Ltd.*, a developer of fanjet engines, sued the defendant, a designer, for breach of contract over a piece of engine the defendant designed, which plaintiff claimed was "worthless." 326 F.3d 1333, 1343–45 (11th Cir. 2003). The court found that a defense expert in the field of aerospace engineering, proffered to provide support that the piece worked as described, was properly admitted. *Id.* The court also found the error rate to be "relatively low." *Id.* Plaintiff argued that the expert had entered data incorrectly, but the court held that the mistake would not affect the validity of the evidence. *Id.*

36. *See, e.g.*, United States v. Moreland, 437 F.3d 424, 427–28, 430–31 (4th Cir. 2006) (upholding the decision to admit testimony of a forensic chemist expert, because although she could not identify an error rate, reviews of her work found it to be error-free); United States v. Vitek Supply Corp., 144 F.3d 476, 480, 485–86 (7th Cir. 1998) (finding that the rate of error was accounted for in the test's control over false positives and false negatives in plaintiffs' proffered scientific evidence, showing the presence of a known carcinogen in defendant's livestock feed used to reduce fat in animals); United States v. Beasley, 102 F.3d 1440, 1444, 1446–48 (8th Cir. 1996) (finding that DNA evidence was properly admitted, as it was tested, reviewed, and accepted by the scientific community, and in regards to the error rate, finding that the test had followed acceptable standards and, therefore, the accuracy of the test was statistically significant); United States v. Chischilly, 30 F.3d 1144, 1148, 1152, 1154–55 (9th Cir. 1994) (holding that although the rate of error was not quantified, there was a sufficient showing in support of a low error rate to allow expert testimony matching the defendant's DNA with that found on the body of a sexual abuse victim, and noting, specifically, the minimal existence of false positives); Wright v. Case Corp., No. 1:03-CV-1618-JEC, 2006 U.S. Dist. LEXIS 7683, at *14 (N.D. Ga. Feb. 1, 2006) (granting defendant's motion to exclude plaintiff's mechanical engineering expert, because the expert's alternative designs for the seat safety bar were not reliable due to potential feasibility issues, and because the associated error rate was therefore unquantifiable but potentially very high).

37. Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 594 (1993) (citations omitted).

38. Daubert v. Merrell Dow Pharm., Inc., 43 F.3d 1311, 1317 n.4 (9th Cir. 1995). On remand, the Ninth Circuit affirmed the district court's ruling that the plaintiff's experts, providing opinions that a prescription drug taken during pregnancy caused birth defects in infants, were inadmissible, leading to a summary judgment in favor of the defendant. *Id.* at 1322. In applying the *Daubert* factors, the court found that interpretation of the third factor—rate of error—was particularly difficult, primarily because two of the three experts did not actually test their theories, making an error rate impossible to define. *Id.* at 1317 n.4.

> otherwise, if (1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case.[39]

This second attempt at the complex business of codifying the Federal Rules of Evidence to allow opinion testimony suggests at least two things: first, that the language of Rule 702 needs to be redrafted to use clearer language and to be more consistent with existing authority; and second, that the prevailing four factors of *Daubert* are insufficient for guiding determinations of admissibility.[40] Nevertheless, the *Daubert* factors continue to assert a powerful force in the admissibility of expert evidence.

II. Error in the Courts

In Part II, we empirically review how courts have handled the "known or potential rate of error" ("KPRE") factor[41] at trial and on appeal. To generally understand how the courts have managed the KPRE factor, we collected all federal trial and federal appellate cases prior to October 2008 that satisfied the criteria of citing *Daubert* and, specifically, using either the term "error rate" ("ER") or "known or potential rate of error."[42]

A. Trial Court Results

Our research produced 1585 trial cases with several thousand experts. Using stratified random sampling with different circuits as the strata, we identified 107 trial cases for our research.[43] Taken together, these cases provided 200 experts who were challenged for admissibility.[44]

---

39. Fed. R. Evid. 702.

40. *See* Fed R. Evid. 702 advisory committee's note.

41. *Daubert*, 509 U.S. at 594.

42. The search concluded in October 2008. To develop a dataset of cases from which we could draw a sample, we used the LexisNexis online database. Our search was repeated three times per circuit: First by a key term search of "known or potential rate of error," then by a key term search of "error rate," then by both "known or potential rate of error" and "error rate" (to identify any cases using both phrases and, therefore, listed twice within the collected results).

43. Although stratified sampling ordinarily is designed to balance the sample—in other words, to achieve a constant proportion—across each stratum, our method was primarily designed to avoid lack of representation in any one of the circuit courts of appeal. *See* Gary L. Tietjen, A Topical Dictionary of Statistics 145 (1986).

44. *See infra* Table 1.

TABLE 1: TRIAL CASES BY CIRCUITS THAT CITE *DAUBERT* AND REFER TO ER OR KPRE[45]

Federal Circuit	Identified	Stratified Random Sampling
1st	58	4
2d	190	12
3d	189	12
4th	68	4
5th	196	12
6th	140	9
7th	172	11
8th	123	8
9th	135	8
10th	150	9
11th	148	9
DC	16	9
Total	**1585**	**107**

At the trial court level, we examined several elements, including the frequency with which experts were scrutinized[46] under each factor and the frequency that the expert was admitted.[47] For example, only 33 of 200 randomly selected experts were analyzed on KPRE, and none of those 33 experts were analyzed on KPRE alone. Not surprisingly, KPRE was the factor considered least among the four. Although the random sample size is insufficient to identify a significant difference in admit proportions among the four factors, KPRE exhibits the second lowest admit proportion. This suggests that perhaps it may be a factor of "last" resort when trial courts prepare their support for not admitting an expert.

45. Note that 107 cases produced 200 experts.

46. We required that the factor be applied to the context of the case. If a factor was mentioned within a case in regards only to a discussion of *Daubert*, rather than in application to the case, we regarded the factor as unaddressed.

47. *See infra* Table 2.

TABLE 2: TRIAL CASES INVOLVING THE ADMISSIBILITY OF EXPERTS CLASSIFIED BY FACTOR[48]

	KPRE		Testability		Peer Review (PR)		General Acceptance (GA)		Total
	Only KPRE	With Other Factors	Only Test-ability	With Other Factors	Only PR	With Other Factors	Only GA	With Other Factors	
Number	0	33	82	69	3	44	5	49	200
Admit	0	16	53	37	2	19	5	26	130
Admit Proportion	N/A	0.48	0.65	0.54	0.67	0.43	1.00	0.53	0.65
95% CI for Admit Proportion	N/A	0.31–0.66	0.54–0.75	0.42–0.65	0.10–1.00	0.29–0.58	0.56–1.00	0.39–0.67	0.58–0.72

Not surprisingly, testability is a favored factor, probably because it is grounded in principles other than science and can be affirmatively demonstrated by simply providing an explanation of how a theory may be tested and, by extension, rejected.[49] General acceptance requires more subjectivity as experts may reasonably disagree as to what is generally accepted within their fields. General acceptance also has problems standing on its own due to the Court's setting it aside in *Frye* as a lone factor.[50] General acceptance, however, has enjoyed the position of sole criterion on admissibility for seventy years,[51] and it is not surprising to see its prevalence in many of the decisions. Peer review is more difficult to assess and thus, less frequently analyzed on its own. This is probably due to the fact that adequate peer review falls squarely within the prevailing culture of science, a realm which the gatekeeper may avoid in favor of more accessible factors. Nevertheless, the data suggest peer review is an easier factor to assess than error.

Also significant in Table 2 is the admissibility frequency: When multiple factors are considered, the likelihood of admission decreases in all cases. We believe that these data suggest, even if only slightly, that trial courts' presumption is toward admission. When admission is rejected, the courts are more careful to provide additional bases. One thing is certain: The covariate[52] of the factor is essential to the analysis.[53]

48. Note that the totals do not add across the rows because of double counting in columns 3, 5, 7, and 9.

49. *See* Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 593 (1993) (citing KARL POPPER, CONJECTURES AND REFUTATIONS: THE GROWTH OF SCIENTIFIC KNOWLEDGE 37 (5th ed. 1989)).

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

51. *See Daubert*, 509 U.S. at 585.

52. A covariate is a relevant but often missing variable in an analysis. JUDEA PEARL, CAUSALITY 78 (2000). In this case, when the specific factor is present in Table 2, the effect is evident. When the effect is obscured through aggregation of the data in Table 3, the effect vanishes. A fine example of a missing covariate is provided in David H. Kaye & David A Freedman, *Reference Guide on Statistics*, *in*

When ignoring the covariate of factor, the effect essentially vanishes. Note the admit proportion is 64% when ignoring specific factors but becomes 48%, 54%, 43%, and 53%, respectively, when considered by factor. This is an example of a pseudo-Simpson's Paradox.[54]

TABLE 3: TRIAL CASES INVOLVING THE ADMISSIBILITY OF EXPERTS CLASSIFIED BY SINGLE FACTOR AND MULTIPLE FACTORS

	Single Factor Only	**Multiple Factors**	**Total**
Number	90	110	200
Admit	60	70	130
Proportion	0.67	0.64	0.65

Concerning the covariate of the factor, we considered whether each factor was a significant predictor of admissibility. Using the entire random sample of 200 experts, we created a simple two-by-two contingency table[55] for each factor. One variable was whether the factor was considered or not in the decision whether to admit the expert. The second variable was whether the expert was admitted or not.[56] It is evident from the *p*-values that all factors were associated with the decision to admit except for general acceptance, which was not statistically significantly linked.[57] More specifically, when the statistically significant factors were considered, admit proportions dropped. Whenever KPRE,[58] peer review[59] or testability[60] were analyzed, apart or together, there was a statistically significant drop in the likelihood of admission when compared to all other cases.

REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 83, 108–10 & n.10 (Fed. Judicial Ctr. Ed., 2d ed. 2000), in which the missing covariate is the college attended by the people in the sample.

53. *See infra* Table 3.

54. Simpson's Paradox concerns the contradictory result that occasionally occurs when a covariate is missing. PEARL, *supra* note 52. A fine example of a Simpson's Paradox is evident within the missing covariate example provided in Kaye & Freedman, *supra* note 52, at 108–10.

55. *See generally* DAVID J. SHESKIN, HANDBOOK OF PARAMETRIC AND NONPARAMETRIC STATISTICAL PROCEDURES 209–11 (1997) (discussing *r*X*c* contingency tables).

56. *See infra* Tables 4, 5, 6, & 7.

57. *See infra* Table 7.

58. *See infra* Table 4.

59. *See infra* Table 5.

60. *See infra* Table 6.

Table 4: Cross-Tabulation of Experts by Admission and by Consideration of KPRE[61]

	Court Did Consider KPRE	Court Did NOT Consider KPRE	Total
Admit	16	114	130
NOT Admit	17	53	70
Total	33	167	200
Proportion	0.48	0.68	0.65

Table 5: Cross-Tabulation of Experts by Admission and by Consideration of PR[62]

	Court Did Consider PR	Court Did NOT Consider PR	Total
Admit	21	109	130
NOT Admit	26	44	70
Total	47	153	200
Proportion	0.45	0.71	0.65

Table 6: Cross-Tabulation of Experts by Admission and by Consideration of Testability[63]

	Court Did Consider Testability	Court Did NOT Consider Testability	Total
Admit	90	40	130
NOT Admit	61	9	70
Total	151	49	200
Proportion	0.60	0.82	0.65

Table 7: Cross-Tabulation of Experts by Admission and by Consideration of GA[64]

	Court Did Consider GA	Court Did NOT Consider GA	Total
Admit	31	99	130
NOT Admit	23	47	70
Total	54	146	200
Proportion	0.57	0.68	0.65

61. Chi Square = 4.74, $p \approx 0.0295$.
62. Chi Square = 11.15, $p \approx 0.0008$.
63. Chi Square = 7.89, $p \approx 0.0050$.
64. Chi Square = 1.87, $p \approx 0.1710$.

Taken together, our findings from the randomly selected trial cases suggest that trial courts tend to avoid KPRE (and to a lesser degree, peer review). Whenever trial courts considered these two factors, they tended to use the factors to support their decisions not to admit the expert.

B. Appellate Court Results

Applying the same search criteria used to identify trial cases, we identified 337 experts considered in 271 appellate cases satisfying our search criteria. These data are a complete census of our findings at the appellate level, unlike the trial cases in which the large numbers of identified cases and experts necessitated sampling.[65] At the appellate level, we were primarily interested in whether trial decisions were reversed or not and whether the appellate court identified KPRE in its reasoning.

Table 8: Experts Identified in Appellate Cases: Cases That Refer to ER or KPRE

Fed Circuit	Experts Identified	Number of Experts or Judges that Directly Addressed ER or KPRE	Proportion of Experts or Judges Directly Addressing ER or KPRE	Reversals	Reversals with ER or KPRE Addressed in the Opinion
1st	12	3	0.25	1	0
2d	11	0	0.00	1	0
3d	19	3	0.16	6	2
4th	22	3	0.14	0	0
5th	36	1	0.03	12	0
6th	32	2	0.06	5	0
7th	25	2	0.08	7	0
8th	33	9	0.27	11	3
9th	52	11	0.21	10	0
10th	46	2	0.04	4	0
11th	40	4	0.10	8	0
DC	9	1	0.11	3	0
Totals	**337**	**41**	**0.12**	**68**	**5**

Evident from Tables 9 and 10 below is the finding that reversals drop when ER or KPRE is directly addressed in the lower court. Although not persuasive standing alone, it is additional evidence that the KPRE factor may be troublesome. When appellate courts review cases that directly address the KPRE factor for experts not admitted, the reversals drop from 15% to 11% ($p = 0.28$). Similarly, when appellate courts review cases that directly address the KPRE factor for experts admitted, the reversals drop from 25% to 14% ($p = 0.08$). These results, taken together, could indicate one of at least two things about the appellate courts: (1) they are less comfortable with the substantive issues of KPRE and thus more deferential, but nevertheless feel compelled to

65. *See infra* Table 8.

consider it as a matter of law, or (2) they find that the substantive issues of KPRE provide an improved measure of objectivity in the gatekeeper's decision and, accordingly, are more deferential. In either case, these data and inferences support the idea that the KPRE factor would benefit from a carefully constructed definition that would be easy to implement but nevertheless confer objectivity upon the decision to admit expert evidence.

TABLE 9: EXPERTS ($n = 337$) WITHIN APPELLATE CASES THAT REFER TO ER OR KPRE[66]

		Reversed		Proportion Reversed
		N	**Y**	
Admitted	**N**	141	25	0.15
	Y	128	43	0.25

TABLE 10: EXPERTS ($n = 41$) WITHIN APPELLATE CASES: EXPERTS OR JUDGES THAT DIRECTLY ADDRESS ER OR KPRE[67]

		Reversed		Proportion Reversed
		N	**Y**	
Admitted	**N**	17	2	0.11
	Y	19	3	0.14

C. SUMMARY OF RESULTS

Our empirical analyses confirm our hypothesis: *Daubert* is difficult to implement. More specifically, the nonexclusive factor of KPRE is especially problematic. Its prevalence among the four enumerated factors is the lowest. In fact, within a relatively large representative sample of cases, we never found it to be the sole criterion for admissibility. We did, however, find each of the other factors serving as sole criterion in a number of cases. Whenever KPRE was applied in conjunction with other factors, its application was inconsistent with scientists' concept of error.

In our analyses of appellate decisions, we found that only 12% of the experts or judges in these decisions have directly dealt with KPRE. Furthermore, when the appellate court reversed the trial court on the admissibility of expert opinion—in 20% of the cases—the appellate court relied on KPRE in only 7% of those reversals—the entirety of which (5 of 68) came from only two circuits. Our scientific training, together with these data, leads us to believe that the KPRE factor ought to be the essential factor of admissibility. As it is the most problematic, proper

66. Chi Square = 5.32, $p \approx 0.0211$.
67. Chi Square = 0.09, $p \approx 0.7615$.

consideration of this factor demands attention.

III. A Taxonomy of Error

Most people can find some comfort in the quantification of error, if for no other reason than that correct quantification provides objectivity. Scientists and professionals regularly assert confidence intervals[68] and *p*-values to make their points.[69] There is something satisfying in summarily reducing the research into a simple decision rule that requires only elementary math to appreciate. Confidence intervals and *p*-values, however, only assess the random error evident in the research.[70] Because great learning and effort is undertaken by the scientist in creating this summary—indeed this is one of the greatest parts of a scientist's training—there is something of an "endowment effect" on the process and its findings.[71] Researcher and reader alike come to appreciate the effort, the learning, and the data, and are eager to engage this scientific currency of *p*-values and confidence intervals to assess a scientific endeavor's value.[72]

Data derived from scientific and professional inquiry entails three specific types of error. These errors are known by other names, but are most conveniently referred to as measurement error, random error, and systematic error.[73]

Table 11: Types of Uncertainty

Type of Uncertainty	**Synonyms**	**Antonyms**	**Basic Cure in Most Cases**
Measurement Error	Measurement Variation, Mistake, Fraud	Precision	Proper Equipment and Proper Procedures Lead to Greater Precision
Random Error	Variation, Chance, Noise	Reliability, Consistency	Larger Sample Sizes Lead to Greater Reliability
Systematic Error	Systematic Variation, Bias	Validity, Accuracy	Random Sampling or Randomization Enhances Validity

68. Stacie Ezelle Taylor, *Confidence Intervals*, *in* 1 Encyclopedia of Epidemiology 225–27 (Sarah Boslaugh ed., 2008).

69. Mark Gerald Haug, *p-Value*, *in* 2 Encyclopedia of Epidemiology 863–65 (Sarah Boslaugh ed., 2008).

70. Kaye & Freedman, *supra* note 52, at 161, 168.

71. John R. Nofsinger, The Psychology of Investing 36 (2d ed. 2005).

72. The first Author, in his law practice, has witnessed this phenomena—summarily reducing KPRE to a question of *p*-value or confidence interval—among other attorneys and judges in toxic tort litigation and class actions, as well as in prominent Continuing Legal Education (CLE) venues.

73. *See* Rebecca Harrington & Li-Ching Lee, *Type I and Type II Errors*, *in* 2 Encyclopedia of Epidemiology, *supra* note 69, at 1052, 1052–53; David L. McArthur, *Measurement*, *in* 2 Encyclopedia of Epidemiology, *supra* note 69, at 645, 645; Louise-Anne McNutt et al., *Bias*, *in* 1 Encyclopedia of Epidemiology, *supra* note 68, at 77, 77.

A. Measurement Error

Measurement error concerns the uncertainty in measuring a fixed quantity.[74] To measure a mile is a trivial task—a fixed quantity measured out with whatever tool is handy. Note, though, that the tool may not be perfectly measured out to exactly one mile. Also, the operator measuring the mile may err in application of the tool.[75] In a scientific endeavor where the stakes may be sufficiently high—unlike the mile measurement above—these potential measurement errors are controlled with sufficient care relative to the theory studied and its implications.[76]

B. Random Error

Random error concerns chance: the reality that uncertainty exists due to matters beyond our perception.[77] For example, the stock market exhibits random error—if it did not, the efficient market naysayers would seize upon every change and become extraordinarily wealthy. A more mundane and illustrative example is the simple die. Its sole purpose is to introduce random variation. Random error occurs when we reasonably expect the "expected value" of 3.5, but never actually observe such a value.[78] More relevant to legal practice, random error is contemplated in employment law's theory of disparate impact,[79] where a plaintiff's prima facie case may be made on a *p*-value of around 0.05.[80] The *p*-value is nothing more than a statement of random error.

C. Systematic Error

We believe that an expert's account of systematic error is far more important than that of random error. While both are important, systematic error goes to the question of relevance: to what degree does the research generalize to the person(s) in question?[81] Random error merely is an artifact of research protocol—namely, sample size: given

---

74. *See* McArthur, *supra* note 73, at 645–49.

75. Each of these concerns may be considered as sub-units of measurement error. Many scientific disciplines have their own nomenclature, but in the end, these ideas are collectively measurement error.

76. *See* Mark Haug & Mark Hirschey, *The January Effect*, 62 Fin. Analysts J., Sept./Oct. 2006, at 78, 79 (arguing that by examining small-cap stocks instead of more profitable trading opportunities, the effect being studied is more likely a statistical oddity than compelling evidence of market inefficiency).

77. *See* Harrington & Lee, *supra* note 73, at 1053 ("[W]e accept the probability of drawing incorrect conclusions merely by chance.").

78. *See generally* Deborah J. Bennett, Randomness (1998).

79. *See, e.g.*, Teamsters v. United States, 431 U.S. 324, 339 (1977) (considering statistical evidence of employment discrimination).

80. 1 David L. Faigman et al., Modern Scientific Evidence: The Law and Science of Expert Testimony § 6:36, at 360 n.3, 362 (2008–2009 ed. 2008).

81. Kaye & Freedman, *supra* note 52, at 160.

that to which we can generalize, to what degree is the picture in focus?[82] We believe it unfortunate that footnote 9 ever found its way into *Daubert*—as Starrs's point cited therein is incorrect.[83] Scientific reliability is a question of random error, while scientific validity is a question of systematic error.[84]

Systematic error, more commonly known as bias, is a great fact of the scientific endeavor.[85] A large volume of litigation involves human injury such as toxic torts, medical malpractice, and products liability. In such matters, proximate cause is typically established with scientific evidence demonstrating an association between the action complained of and the plaintiff's injury or damages.[86] Unfortunately, such scientific findings of association often do not enjoy the benefits of an experimental protocol, but rather an observational protocol.[87] For example, animal studies are extremely useful in elevating human subject research from observational to experimental—permitting random sampling and its accompanying benefits.[88] Criticism, however, often draws attention to the differences between the animal and the human—the sole source of bias in a well-designed animal study—while largely ignoring the many biases attendant to the alternative.[89]

The distinction between experimental (elements of random sampling[90] or randomization[91]) and observational (no random

---

82. *See generally* BENNETT, *supra* note 78.

83. Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 590 n.9 (1993) ("[T]he difference between accuracy, validity, and reliability may be such that each is distinct from the other by no more than a hen's kick." (quoting James E. Starrs, Frye v. United States *Restructured and Revitalized: A Proposal to Amend Federal Evidence Rule 702*, 26 JURIMETRICS J. 249, 256 (1986))).

84. *See* McArthur, *supra* note 73.

85. *See* McNutt et al., *supra* note 73.

86. *See* Michael D. Green et al., *Reference Guide on Epidemiology*, *in* REFERENCE MANUAL ON SCIENTIFIC EVIDENCE, *supra* note 52, at 333, 374 ("Once an association has been found between exposure to an agent and development of a disease, researchers consider whether the association reflects a true cause-effect relationship.").

87. An experimental protocol contains some element of random sampling or randomization that eliminates experimental bias beyond the point of the randomizing/random sampling. *Id.* at 391. For example, clinical trials enjoy the benefits of an experimental protocol. *See* John J. Hsieh, *Clinical Trials*, *in* 1 ENCYCLOPEDIA OF EPIDEMIOLOGY, *supra* note 69, at 201–02. An observational protocol has no element of either randomization or random sampling. Green et al., *supra* note 86, at 394. Accordingly, scientists must thoroughly consider their research designs to minimize biases to the extent possible. *See* Craig Newschaffer, *Observational Studies*, *in* 2 ENCYCLOPEDIA OF EPIDEMIOLOGY, *supra* note 69, at 757–60.

88. Bernard D. Goldstein & Mary Sue Henifin, *Reference Guide on Toxicology*, *in* REFERENCE MANUAL ON SCIENTIFIC EVIDENCE, *supra* note 52, at 405.

89. *Id.* at 413–14. David Sackett has catalogued thirty-five biases that may accompany the sort of scientific studies that would be appropriate in the absence of animal studies. *See generally* David L. Sackett, *Bias in Analytic Research*, 32 J. CHRONIC DISEASES 51 (1979).

90. *See* Anthony Roman, *Sampling Techniques*, *in* 2 ENCYCLOPEDIA OF EPIDEMIOLOGY, *supra* note 69, at 935–38.

91. *See* Sydney Pettygrove, *Randomization*, *in* 2 ENCYCLOPEDIA OF EPIDEMIOLOGY, *supra* note 69, at 891.

component) derives from the same features that systematic error does. For example, a toxicology study may use humans as it subjects. Of course, there ought to be no random component—as that would inevitably place someone in harm's way against his wishes. Using an observational study, for instance, a cohort study,[92] we could track people over time and study the relationship between those who were exposed to the suspected agent and the suspected outcome(s). If a relationship exists, we are wary of the relationship, as there may be other factors not studied that are the true causes of the outcome.[93] Random sampling and randomization solves this problem by creating statistically equivalent groups.[94] Therefore, a scientist may prefer rats as subjects. When all is said and done, rat studies enjoy the benefits and expediency of experimental studies. Of course, there now remains a new source of systematic bias: Rats are systematically different than humans. A proper scientific analysis of either protocol would identify and enumerate the sources of systematic error.

Conclusion

The idea of error summarily captures the whole of an expert's credibility. If evidence is without error, then we welcome it, except where it may be unjust to do so. If evidence is with error—and that error can be properly accounted for—then we would also welcome it as evidence, except where it may be unjust to do so, as it would "assist the trier of fact to understand the evidence or to determine a fact in issue."[95] Our proposal is aimed toward this end:

> If an expert can account for the measurement error, the random error, and the systematic error in his evidence, then he ought to be permitted to testify. On the other hand, if he should fail to account for any one or more of these three types of error, then his testimony ought not be admitted.

Our proposal satisfies all four criteria outlined in *Daubert*, is more expedient than *Daubert* or Federal Rule of Evidence 702, is more manageable to the gatekeeper, and will result in better decisions through its reliance on objectivity. Where there is disagreement as to its elements, "vigorous cross-examination" ensures better decisions.

Certainly an accounting of error more fully responds to KPRE than does any other alternative—which usually entails an analysis of the

92. *See* Philip C. Nasca, *Study Design*, *in* 2 Encyclopedia of Epidemiology, *supra* note 69, at 1008, 1010–11.

93. *See* Green, *supra* note 86, at 369 ("Even when an association exists, researchers must determine whether the exposure causes the disease or whether the exposure and disease are caused by some other confounding factor.").

94. *See supra* notes 90 and 91.

95. Fed. R. Evid. 702.

"inferior" random error.[96] A full accounting of error also satisfies testability, as such an accounting can only be derived from empirical evidence. To have empirical evidence is to have testability, as testing is the source of empirical evidence. Peer review and general acceptance find their way into the "*Daubert* calculus" through judicial deference towards the expert's profession: "What do other experts in this field make of this evidence?"

Nearly every scientist's academic training includes coursework in statistics either in name or through other coursework concerning research methodology. What all these courses have in common is a thorough treatment of random error. Often in the social sciences, they include a thorough treatment of measurement error and in all the sciences, admonishment to minimize measurement error and systematic error to the extent possible in any given research. Unquestionably, peer review and general acceptance, within any scientific discipline, rest squarely on this cornerstone of science.

Our proposal is also more expedient than *Daubert* or Federal Rule of Evidence 702. *Daubert*'s four nonexclusive factors and requirement of "relevance" constitute an important attempt to achieve just resolutions in litigation. Similarly, amended Federal Rule of Evidence 702's three criteria are "codified" positions to build upon *Daubert*'s standard for admissibility. Each attempt seeks to insure that scientific evidence be admitted when it is objective and untethered to advocacy, and excluded when it is not. We propose an abandonment of these positions and encourage the courts to deal with the interest squarely: to nail down objectivity directly at its enabling source: the accounting of error.

When an expert is required to account for error, litigation is much more manageable to the gatekeeper. So-called "*Daubert* hearings" properly place the burden of admissibility on the litigants.[97] Current hearings, which often consist of confusing arguments—one side borrowing some elements, while the other side borrowing other elements—easily become contests of covert obfuscation (resembling what every parent of two or more children knows all too well), entreating the gatekeeper *cum* parent. Our proposal, on the other hand, anchors the hearings squarely on three definitive criteria, all of which must be addressed. There is nothing more or anything less for the judge to

96. Random error is inferior to systematic error by virtue of its origin. Random error derives from sample size—smaller sample sizes yield more random error—and alternatively, larger samples convey consistency (scientific reliability). Kaye & Freedman, *supra* note 52, at 171. Random error says nothing as to whether the results are "on point"—the "fit" of the evidence to the case at bar. Systematic error, on the other hand, derives from lack of fit between the "inequality" of the evidence and the case at bar—ordinarily sampling that is not random. In animal studies, however, the source of systematic error stems from the differences between the animal and the human. *Id.* at 160.

97. *See, e.g.*, Bourjaily v. United States, 483 U.S. 171, 176 (1987).

consider. Can the expert account for the three errors? Can the opponent adequately dispute the accounting? At a properly adversarial hearing, such directed purpose and clear rules will assist nearly any jurist in determining whether the proposed expert can adequately account for error. Such a hearing would free the judge of having to rely on notions or presumptions to distinguish science from mere "junk science."[98]

Moreover, our proposal will result in better decisions through its reliance on objectivity. The admissibility of expert evidence historically has been difficult to address properly[99] due to its high impact on outcomes in litigation, as well as the high potential for courts to decide admissibility inconsistently. It is precisely because expert evidence has such a great influence on outcomes, and rightfully so where experts can cast clarifying light on a matter, that we ought to desire consistent application of the rules of admissibility. By writing the rules around the three sources of error, we achieve the interest of objectivity, we obtain consistency in the application of the rules, and we maintain deference to the scientific endeavor. These three outcomes, taken together, will result in better decisions, and ultimately the justice[100] that they serve.

---

98. *See* KENNETH R. FOSTER & PETER W. HUBER, JUDGING SCIENCE: SCIENTIFIC KNOWLEDGE AND THE FEDERAL COURTS 16–17 (1999).

99. *See* GOLAN, *supra* note 16, at 5–7.

100. "Justice" is an unfortunate victim of the inevitable clutter of language amidst life and conflict. Contemporary usage has relegated "justice" to merely an act of the will, or less, an act of emotion, rather than to its ancient place as the second of the four cardinal virtues. *See generally* JOSEF PIEPER, THE FOUR CARDINAL VIRTUES (1966). It is in the spirit of the cardinal virtues, namely, the cardinal virtue of prudence (an even more tortured victim of contemporary usage) and of justice, that motivates our proposal. *See generally id.*