

Eyes Wide Shut: Hidden Problems and Future Consequences of the Fact-Based Validity Standard

Joëlle Anne Moreno*

Over the past decade *Daubert*¹ has transformed judicial decisions involving questions of science and law. *Daubert* governs the federal courts and the 30 states that adopted *Daubert* in whole or in part.²

* Associate Professor, New England School of Law. I would like to thank Professors D. Michael Risinger and Mark P. Denbeaux for inviting me to this excellent conference and Professors Ronald J. Allen and Paul C. Giannelli for their provocative and insightful comments.

¹ *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579 (1993).

² See *Alabama—S. Energy Homes, Inc. v. Washington*, 774 So. 2d 505, 516-17 (Ala. 2000) (acknowledging that the legislature has used *Daubert* with respect to DNA evidence, but not explicitly switching standard from *Frye* to *Daubert* for other evidence); *Alaska—State v. Coon*, 974 P.2d 386, 388-99 (Alaska 1999) (adopting the *Daubert* standard); *Arizona—Logerquist v. McVey*, No. CV-98-0587-PR, 2000 WL 419980 (Ariz. Apr. 19, 2000) (en banc) (retaining the *Frye* standard); *Arkansas—Moore v. State*, 915 S.W.2d 284, 293-94 (Ark. 1996) (recognizing the *Daubert* standard but not expressly adopting it); *California—People v. Leahy*, 882 P.2d 321, 323-24 (Cal. 1994) (refusing to adopt *Daubert* and noting that California has long held to the *Frye* standard and would continue to do so); *Colorado—Brooks v. People*, 975 P.2d 1105, 1113 (Colo. 1999) (en banc) (noting that Colorado "has neither explicitly endorsed nor rejected the *Daubert* analysis"); *Connecticut—State v. Porter*, 698 A.2d 739, 751 (Conn. 1997) (adopting the *Daubert* standard); *Delaware—Bell Sports, Inc. v. Yarusso*, 759 A.2d 582 (Del. 2000) (expressly adopting *Daubert*); *Florida—Brim v. State*, 695 So. 2d 268, 271-72 (Fla. 1997) (rejecting *Daubert*); *Georgia—Jordan v. Ga. Power Co.*, 466 S.E.2d 601, 604-05 (Ga. 1995) (applying state law and not adopting *Daubert*); *Hawaii—State v. Fukusaku*, 946 P.2d 32, 42 (Haw. 1997) (refusing to follow *Daubert*); *Idaho—State v. Trevino*, 980 P.2d 552, 557-58 (Idaho 1999) (adopting the *Daubert* standard); *Indiana—Sears Roebuck & Co. v. Manuilov*, 742 N.E.2d 453, 462 (Ind. 2001) (retaining the *Frye* standard); *Iowa—Leaf v. Goodyear Tire & Rubber Co.*, 590 N.W.2d 525, 530-33 (Iowa 1999) (adopting a limited application of *Daubert*); *Kansas—State v. Canaan*, 964 P.2d 681, 691-92, 694 (Kan. 1998) (retaining the *Frye* standard); *Kentucky—Mitchell v. Commonwealth*, 908 S.W.2d 100 (Ky. 1995) (*Daubert* expressly adopted); *Louisiana—State v. Ledet*, 792 So. 2d 160 (La. 2001) (adopting the *Daubert* standard); *Maine—State v. McDonald*, 718 A.2d 195 (Me. 1998) (adopting *Daubert*); *Maryland—Hutton v. State*, 663 A.2d 1289, 1295-96 n.10 (Md. 1995) (determining that Maryland will still follow the *Frye* standard despite the fact that Maryland's Rules of Evidence are patterned after the Federal Rules of Evidence and were passed into legislation after the *Daubert* decision); *Massachusetts—Commonwealth v. Senior*, 744 N.E.2d 614 (Mass. 2001) (applying various *Daubert* factors); *Michigan—Although the Michigan Supreme Court has not addressed the issue, the consensus among the lower courts favors Frye. See, e.g. People v. Coy*, No. 238112, 2003 Mich. App. LEXIS 1880, at *12 n.3 (Mich. Ct. App.

Even in states that retain a *Frye*-type “general acceptance” standard,³ most state judges report that *Daubert* exerts a powerful influence on their admissibility decisions.⁴ Since *Daubert*, judges must decide for

Aug. 7, 2003) (stating that until the Michigan Supreme Court holds otherwise, the lower courts are bound to apply *Frye*). The Michigan legislature, however, has codified *Daubert* for expert scientific opinion in actions for “death of a person or for injury to a person or property” See MICH. STAT. ANN. 600.2955 (Michie 2003); Minnesota—State v. Klawitter, 518 N.W.2d 577, 585 & n.3 (Minn. 1994) (noting that the *Frye* standard has been utilized before and after *Daubert* although expressing that “we do not address the effect of the *Daubert* decision on the use or application of the *Frye* rule in Minnesota”); Mississippi—Gleaton v. State, 716 So. 2d 1083, 1087 (Miss. 1998) (retaining the *Frye* standard); Missouri—Callahan v. Cardinal Glennon Hosp., 863 S.W.2d 852 (Mo. 1993) (continuing to apply *Frye*); Montana—State v. Moore, 885 P.2d 457 (Mont. 1994) (adopting the *Daubert* standard); Nebraska—Sheridan v. Catering Mgmt., Inc., 566 N.W.2d 110, 113 (Neb. 1997) (retaining the *Frye* standard); Nevada—Krause Inc. v. Little, 34 P.3d 556, 569 (Nev. 2001) (reiterating an earlier rejection of *Daubert*); New Hampshire—State v. Cort, 766 A.2d 260 (N.H. 2000) (applying various *Daubert* factors); New Jersey—State v. Harvey, 699 A.2d 596, 621 (N.J. 1997) (applying various *Daubert* factors); New Mexico—State v. Anderson, 881 P.2d 29 (N.M. 1994) (adopting the *Daubert* standard); New York—People v. Wernick, 674 N.E.2d 322, 324 (N.Y. 1996) (retaining the *Frye* standard); North Carolina—State v. Goode, 461 S.E.2d 631, 639, 641 (N.C. 1995) (adopting the *Daubert* standard); North Dakota—City of Fargo v. McLaughlin, 512 N.W.2d 700, 705 n.2 (N.D. 1994) (retaining the *Frye* standard); Ohio—Miller v. Bike Athletic Co., 687 N.E.2d 735 (Ohio 1998) (adopting *Daubert* factors); Oklahoma—Christian v. Gray, 65 P.3d 591 (Okla. 2003) (expressly adopting *Daubert/Kumho* for civil matters and noting that the Oklahoma Court of Criminal Appeals has already done the same for criminal matters); Pennsylvania—Commonwealth v. Arroyo, 723 A.2d 162, 170 & n.10 (Pa. 1999) (retaining the *Frye* standard); Rhode Island—State v. Quattrocchi, 681 A.2d 879, 884 n.2 (R.I. 1996) (adopting the *Daubert* standard); South Carolina—State v. Council, 515 S.E.2d 508, 518 (S.C. 1998) (using factors similar to, but not specifically adopting, the *Daubert* factors); South Dakota—State v. Hofer, 512 N.W.2d 482 (S.D. 1994) (adopting the *Daubert* standard); Tennessee—McDaniel v. CSX Transp., Inc., 955 S.W.2d 257 (Tenn. 1997) (adopting the *Daubert* standard); Texas—E.I. du Pont Nemours & Co. v. Robinson, 923 S.W.2d 549 (Tex. 1995) (adopting the *Daubert* standard); Utah—State v. Butterfield, 27 P.3d 1133 (Utah 2001) (stating that the test for admissibility requires threshold showing of “inherent reliability”); Vermont—State v. Brooks, 643 A.2d 226 (Vt. 1993) (adopting the *Daubert* decision); Virginia—Spencer v. Commonwealth, 393 S.E.2d 609, 621 (Va. 1990) (expressly declining to follow *Frye*, but not adopting *Daubert*); Washington—State v. Copeland, 922 P.2d 1304, 1310 (Wash. 1996) (retaining the *Frye* standard); West Virginia—Wilt v. Buracker, 443 S.E.2d 196 (W. Va. 1993) (adopting the *Daubert* decision); Wisconsin—State v. Peters, 534 N.W.2d 867 (Wis. 1995) (basing admissibility on three-part relevance test); Wyoming—Bunting v. Jamieson, 984 P.2d 467 (Wyo. 1999) (adopting the *Daubert* standard). The District of Columbia has not yet adopted Federal Rule of Evidence 702, and there has been no majority opinion that has addressed *Daubert*. Cf. Taylor v. United States, 661 A.2d 636, 651-52 (D.C. 1995) (Newman, S.J., dissenting) (urging the adoption of Federal Rule of Evidence 702 and *Daubert*).

³ *Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923).

⁴ See Sophia I. Gatowski et al., *Asking the Gatekeepers: A National Survey of Judges on Judging Expert Evidence in a Post-Daubert World*, 25 LAW & HUM. BEHAV. 433, 443 (2001) (describing how 94% of state court judges surveyed find *Daubert* has either “some value” or “a great deal of value” for their decision-making process on questions involving scientific evidence regardless of whether *Daubert* or *Frye* governs

themselves whether proffered evidence is *scientifically* reliable; they may no longer defer to the scientific community.⁵ *Daubert*, the Supreme Court's response to growing concerns that jurors were using "junk science" to decide legal cases,⁶ contains the implicit assumption that forcing judges into uncharted terrain is better than exposing jurors to evidence they could never hope to understand.⁷

In 1998, the *Kumho*⁸ Court expanded the requirement of autonomous judicial validity screening beyond scientific evidence to include technology-based evidence and a potentially infinite undefined class of evidence that relates to "specialized knowledge."⁹ Although Justice Breyer did not explicitly ground his analysis of the plaintiff's tire engineering evidence in the four *Daubert* scientific reliability criteria, i.e., (1) testability/falsifiability; (2) error rates; (3) peer review and publication; and (4) general acceptance,¹⁰ he emphasized that similar factors should govern the evaluation of technical or specialized evidence.¹¹ According to the *Kumho* Court,

admissibility in their jurisdiction).

⁵ Although the *Daubert* Court used the word "reliable" to refer to the quality of the scientific evidence, I have argued elsewhere that this reflects a misunderstanding of this scientific term of art and that "validity" is a more accurate term. See Joëlle Anne Moreno, *Beyond the Polemic Against Junk Science: Navigating the Oceans that Divide Science and Law With Justice Breyer at the Helm*, 81 B.U. L. REV. 1033, 1065-70 (2001) (describing how "reliability" refers only to the reproducibility of data, even if the data are wrong, while "validity" connotes a connection between the theory or conclusions and the empirical world).

⁶ See generally PETER W. HUBER, GALILEO'S REVENGE: JUNK SCIENCE IN THE COURTROOM (1991). There is an ongoing concern, a decade after *Daubert*, that "jurors may also be influenced by scientific evidence that lacks validity If jurors are unable to differentiate high-quality research from "junk" science, then it is likely that their decisions will be influenced by both methodologically sound and methodologically inferior research, which is clearly an undesirable outcome." See Margaret Bull Kovera et al., *Reasoning About Scientific Evidence Effects of Juror Gender and Evidence Quality on Juror Decisions in a Hostile Work Environment Case*, 94 J. APPLIED PSYCHOL. 362 (1999).

⁷ "While meticulous *Daubert* inquiries may bring judges under criticism for donning white coats and making determinations that are outside their field of expertise, the Supreme Court has obviously deemed this less objectionable than dumping a barrage of questionable scientific evidence on a jury" Joseph Sanders et al., *Legal Perceptions of Science and Expert Knowledge*, 8 PSYCHOL., PUB. POL'Y & L. 139, 152 (2002).

⁸ *Kumho Tire Co. v. Carmichael*, 526 U.S. 137 (1999).

⁹ See *Kumho*, 526 U.S. at 141 (holding that the *Daubert* gate-keeping role "applies not only to expert testimony based on 'scientific' knowledge, but also to testimony based on 'technical' and 'other specialized knowledge'"). The *Daubert* decision had been explicitly limited to scientific evidence. See *Daubert*, 509 U.S. at 590 n.8 ("Our discussion is limited to the scientific context because that is the nature of the expertise offered here.").

¹⁰ See *Daubert*, 509 U.S. at 593-97.

¹¹ The *Kumho* Court makes a fairly bold assumption that criteria taken from (and

“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 of [the relevant] discipline.’”¹² *Kumho* also added what may become a fifth criterion: “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.”¹³ Although *Daubert* and the post-*Daubert* cases clearly state that these factors are neither fixed nor definitive,¹⁴ courts often operate the admissibility standard by applying the factors in turn.¹⁵

designed for) the assessment of “hard” empirical science can be useful for evaluating all manner of specialized knowledge. This is where *Daubert* is least likely to fit its own task at hand and where the doctrine has not yet been developed. In his contribution to this Symposium, Professor Allen presages the conflict. “To ‘apply’ the standards of *Daubert* to this vast and forbidding landscape, as *Kumho Tire* directs, seems to result in a significant mismatch between tool and task. . . . [because] *Daubert* was fashioned with normal science in mind and invokes standard criteria of scientific validation, such as controlled studies and the like.” Ronald J. Allen, *Expertise and the Supreme Court: What is the Problem?*, 34 SETON HALL L. REV. 1, 2 (2003).

¹² *Kumho*, 526 U.S. at 149 (quoting *Daubert*, 509 U.S. at 592).

¹³ *Id.* at 152.

¹⁴ In their contribution to this Symposium, Professors Mark P. Denbeaux and D. Michael Risinger express concern that courts frequently overlook the fact that the *Daubert* factors were intended to be flexible, despite the fact that the *Kumho* Court “further emphasizes (by way of reiterating a part of *Daubert* often unfortunately ignored) that the four factors were not each necessary conditions for a proper reliability warrant, nor were other factors foreclosed.” Mark P. Denbeaux & D. Michael Risinger, *Kumho Tire and Expert Reliability: How the Question You Ask Gives the Answer You Get*, 34 SETON HALL L. REV. 15, 32 (2003). Perhaps this concern will become less pressing in the wake of Judge Pollak’s recent and widely-read decision in *United States v. Llera Plaza*, 188 F. Supp. 2d 549 (E.D. Pa. 2002). Judge Pollak devoted considerable time and attention to this very question, repeatedly mentioning the “flexible” nature of the *Daubert* inquiry, *id.* at 562, and explicitly noting that *Kumho* clarified that “*Daubert*’s list of specific factors neither necessarily nor exclusively applies to all experts in every case.” *Id.* at 562. Judge Pollak specifically links this careful reading of *Kumho* to his decision to reconsider his earlier ruling, that prosecution experts would not be permitted to present expert testimony that certain fingerprints matched because the methodology of fingerprint analysis failed to satisfy three of the four *Daubert* factors. See *United States v. Llera Plaza*, 179 F. Supp. 2d 492 (E.D. Pa. 2002). Underlying the entire decision reexamining the fingerprint evidence is Judge Pollak’s thorough consideration of the question of whether the four *Daubert* factors are a “reasonable measure of the reliability of expert testimony.” 188 F. Supp. 2d at 563-75.

¹⁵ See, e.g., *Lauzon v. Senco Prods., Inc.*, 270 F.3d 681, 686-96 (8th Cir. 2001) (providing a detailed analysis of each of the four *Daubert* factors in turn); see also Denbeaux & Risinger, *supra* note 14, at 32 n.70 (“The four factors have too often been leaden deadweights woodenly applied, inert impediments to the development of a sophisticated approach by the courts to belief warrants for scientific evidence.”). In response, Professors Denbeaux and Risinger have developed an alternative inquiry based on the following four components: “(1) [f]raming the case-specific target

I. THE EMERGENCE OF A FACT-BASED VALIDITY STANDARD

Kumho did more than just expand the range of evidence subject to more intensive and structured pretrial judicial scrutiny. I have previously posited that *Kumho* contained a subtle correction of *Daubert*¹⁶ perhaps intended to address five years of perceived judicial reluctance to embrace the enhanced gate keeping role.¹⁷ Whatever the Court's motivation,¹⁸ *Kumho* refocused the admissibility decision so that the fit/relevance prong of the two-part *Daubert* test became more prominent. Justice Breyer emphasized that judges must assess evidentiary reliability, not in some broad or abstract manner, but as the evidence has been applied by this expert to the narrow "task at hand."¹⁹

issue; (2) [f]raming the case-specific claim of expertise; (3) [d]etermining what available information bears on a rational belief warrant in regard to the reliability of the claimed expertise; [and] (4) [d]etermining the proper case-specific legal standard of certainty for such a belief warrant." *Id.* at 33.

¹⁶ The problem with *Daubert* might have been caused by the fact that the Court placed the general reliability/validity step of the *Daubert* inquiry first. This structure could appear to require that judges first determine whether proposed expert testimony is generally considered "scientific knowledge," before exploring its relevance. Judges who read *Daubert* this way would balk at the requirement assumed that they assess a potentially infinite amount of scientific information, most of which would not be relevant to the facts at issue. See Moreno, *supra* note 5, at 1052-55.

¹⁷ This concern is addressed in the Advisory Committee Notes to the 2000 amendments to Federal Rule of Evidence 702, which state "[a] review of the case law after *Daubert* shows that the rejection of expert testimony is the exception rather than the rule." FED. R. EVID. 702 advisory committee's note; see also Erica Beecher-Monas, *Blinded by Science: How Judges Avoid the Science in Scientific Evidence*, 71 TEMP. L. REV. 55, 58 (1998) ("All too often, however, courts continue to evade the science issues. In far too many jurisdictions, judges are turning a blind eye to the science involved in the evidence before them."); David L. Faigman, et al., *How Good is Good Enough?: Expert Evidence Under Daubert and Kumho*, 50 CASE W. RES. L. REV. 645, 665 (2000) ("In the forensic context, courts have long admitted a surfeit of expertise with little or no evaluation of the foundation upon which the opinion rests."); Jay P. Kesan, *Who Knows Where the Time Goes?: A Critical Examination of the Post-Daubert Landscape*, 52 FOOD & DRUG L.J. 225, 239-40 (1997) (reviewing numerous post-*Daubert* cases and concluding that "the quantum of scientific information that must undergird an expert's methodology to render it scientifically valid and admissible under *Daubert* is quite minimal"). However, a recent empirical study of a sample of federal district court opinions published between 1980 and 1999 concluded that after *Daubert*, "judges are more actively screening expert evidence[;] whether they are doing so in ways that produce better outcomes has not been determined." Lloyd Dixon & Brian Gill, *Changes in the Standards for Admitting Expert Evidence in Federal Civil Cases Since the Daubert Decision*, 8 PSYCHOL. PUB. POL'Y & LAW 251, 251 (2002).

¹⁸ Professor Allen suggests another possible motivation that "no matter how well credentialed and conversant in an established field, an expert may still testify to falsehoods." See Allen, *supra* note 11, at 5.

¹⁹ This is also emphasized in the *Kumho* Court's conclusion that "the question before the trial court was specific, not general. The trial court had to decide whether *this particular expert* had sufficient specialized knowledge to assist jurors 'in deciding

[T]he specific issue before the [district] court was not the reasonableness in general of a tire expert's use of a visual and tactile inspection, [but was instead] the reasonableness of using such an approach . . . to draw a conclusion regarding the particular matter to which the expert testimony was directly relevant.²⁰

Various federal courts have read *Kumho* to require evaluation of the *application* of the science, technology, or specialized knowledge to the specific facts at issue²¹ and this shift in emphasis was also reflected in the December 2000 amendments to Rule 702 of the Federal Rules of Evidence.²²

If scientific, technological, and specialized evidence is likely to be admitted or excluded based upon the court's evaluation of how

the particular issues in the case." *Kumho*, 526 U.S. at 156 (emphasis added) (citations omitted). According to Professor Risinger who has also advanced the idea of a new focus on the "task at hand,"

what is clearly not consistent with *Kumho Tire* is any attempt to approach the issue of reliability globally. That is, reliability cannot be judged globally, "as drafted," but only specifically, "as applied." The emphasis on the judgment of reliability as it applies to the individual case, to the "task at hand," runs through the opinion like a river.

D. Michael Risinger, *Defining the "Task at Hand:" Non-Science Forensic Science After Kumho Tire Co. v. Carmichael*, 57 WASH. & LEE L. REV 767, 773 (2000).

²⁰ *Kumho*, 526 U.S. at 153.

²¹ See, e.g., *Seatrax, Inc. v. Sonbeck Int'l, Inc.*, 200 F.3d 358, 372 (5th Cir. 2000) ("Whether Daubert's suggested indicia of reliability apply to any given testimony depends on the nature of the issue at hand, the witness's particular expertise, and the subject of the testimony. It is a fact-specific inquiry.") (internal citation omitted); *United States v. Smithers*, 212 F.3d 306, 315 (7th Cir. 2000) (noting that the *Kumho* Court engaged in a thorough reexamination of the technology relevant to the facts that had been presented to the district court); *United States v. Brumley*, 217 F.3d 905, 911 (7th Cir. 2000) (finding that "[t]he Supreme Court in *Kumho Tire* explained that the *Daubert* 'gatekeeper' factors had to be *adjusted to fit the facts of the particular case at issue*, with the goal of testing the reliability of the expert opinion") (emphasis added); *United States v. Horn*, 185 F. Supp. 2d 530, 554 (D. Md. 2002) ("[J]udges do not determine the reliability of scientific or technical issues in the abstract but rather in the context of deciding a particular dispute.").

²² Federal Rule of Evidence 702 was amended to include the following (shown in italics):

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, *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.*

FED. R. EVID. 702. The amended Rule 702 has also been interpreted to require a relevance-based focus. See, e.g., *Cloud v. Pfizer Inc.*, 198 F. Supp. 2d 1118, 1130 (D. Ariz. 2001) ("[F]ederal judges must exclude proffered scientific evidence under Rule 702 unless they are convinced that it speaks clearly and directly to an issue in dispute in the case . . .").

well an expert can apply her knowledge and methods to a discrete set of facts, the advantages and disadvantages of this more case-specific approach must be explored.

A. *Local Advantages, Global Problems*

I have previously described some of the advantages of a more fact-specific admissibility inquiry.²³ At the local or individual case level, limiting the scope of the pretrial evaluation to those principles and methods actually *applied* by this expert to a given set of facts has at least two advantages. By limiting the scope of the analysis, this inquiry conforms more realistically to most judges' limited understanding of complex scientific, technical or other highly specialized information. It also serves the omnipresent judicial concerns of timeliness and efficiency. Upon further reflection, and in light of more recent empirical data, the global or systemic disadvantages of a fact-intensive validity assessment have become more apparent.

Recent empirical data demonstrate that an overwhelming number of judges do not comprehend the basic scientific principles that underlie the *Daubert* decision. This study concludes that two of the four *Daubert* factors, testability/falsifiability and error rates, are uniformly misunderstood and misapplied.²⁴ Although this same study finds that most judges accurately comprehend and apply the only other *Daubert* criterion added to the *Frye* standard, i.e., peer review and publication, facility with this remaining factor provides cold comfort in light of even more recent biomedical data.

Last year the JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION ("JAMA"), dedicated an entire issue to reporting on extensive recent empirical data that cast grave doubts on the general assumption, adopted wholeheartedly by the *Daubert* Court, that peer review actually enhances the validity of scientific publications.²⁵ The problem with courts using peer review and publication to assess the validity of scientific evidence lies not with judges (who can understand and operate this criterion), but with the scientists themselves. The JAMA data expose hidden impediments that may hinder gate keeping efforts by even the most scientifically sophisticated and conscientious members of the judiciary.

In addition to these troubling new data, we must consider the global or systemic effects of a judicial and legislative movement

²³ See generally Moreno, *supra* note 5.

²⁴ See Gatowski et al., *supra* note 4, at 439.

²⁵ See generally 287 JAMA 2739-2898 (June 5, 2002) [hereinafter JAMA].

towards a more fact-based qualitative assessment of scientific evidence. Case-specific evaluation of science “as applied” runs afoul of bedrock scientific principles, conflicts with widely-accepted methods of gauging validity, and guarantees that we cannot develop consistent or useful precedent.

The danger is that judges who lack scientific facility and perceive their role as limited to an assessment of the science applied by *this* expert in *this* case may be less rigorous in their evaluation. This can be particularly problematic if judges mistakenly assume that an imprimatur of validity from within the discipline itself provides some sort of guarantee. Under these circumstances, it is hard to have faith that over the past decade the *Daubert* admissibility standard has been accurately understood or consistently applied.

II. HIDDEN DANGERS: JUDGES’ PROBLEMS UNDERSTANDING AND APPLYING THE *DAUBERT* CRITERIA OF TESTIFIABILITY/FALSIFIABILITY AND ERROR RATES

A recent study of four hundred judges demonstrates that the vast majority of state court judges cannot comprehend or implement even the most basic scientific concepts.²⁶ The Gatowski study was designed to “assess the level at which the judiciary understand the scientific meaning of the *Daubert* guidelines and how they might apply them when evaluating the admissibility of scientific evidence.”²⁷ It provided information that had never previously been collected, analyzed, or published²⁸ and researchers used a different methodology for data collection. They questioned judges directly, rather than the previously favored method of relying on a retrospective analyses of published judicial opinions.²⁹

While providing important insight regarding the influence of *Daubert*, an empirical analysis of published case law is, by its very nature, restricted to an analysis of post hoc justifications of those writing a decision in a particular case and does not fully capture the judicial decision-making process. Although an empirical analysis of case law provides important data about judges’ normative, case specific reasoning, research has demonstrated that there may be significant differences between published and unpublished cases, and that these differences may be dependent

²⁶ See Gatowski et al., *supra* note 4, at 439. This national survey involved judges throughout the country in states that have adopted *Daubert*, states that have adopted a modified version of *Daubert*, and states that continue to apply *Frye*. See *id.*

²⁷ *Id.* at 438.

²⁸ See *id.* at 433-35.

²⁹ See *id.*

upon the case characteristics analyzed and the legal questions involved.³⁰

The Gatowski study concluded that *Daubert* is neither accurately nor consistently applied.³¹ The overwhelming majority of judges have no real understanding of two of the four *Daubert* criteria.³² Specifically, while 88% of the judges reported that “falsifiability”³³ is a useful guideline for assessing scientific evidence, 96% of these same judges lacked even a basic understanding of this core scientific concept.³⁴ Similarly, 91% of the judges reported that they found “error rates”³⁵ helpful, although when questioned, they had no real understanding of this basic scientific precept.³⁶

These findings lead the researchers to conclude that “judges have difficulty operationalizing the *Daubert* criteria and applying

³⁰ *Id.*

³¹ See Gatowski et al., *supra* note 4, at 443. Other commentators have noted that the Gatowski study substantiated existing “uncertainty about whether judges, despite their best intentions, have the time or training to carry out their gatekeeper responsibilities effectively.” Dixon & Gill, *supra* note 17, at 245.

³² See Gatowski et al., *supra* note 4, at 444-46.

³³ Nine years ago, the *Daubert* Court concluded that falsifiability was “a key question to be answered in determining whether a theory or technique is scientific knowledge that will assist the trier of fact,” and defined falsifiability as “whether it [the theory] can be (and has been) tested.” *Daubert*, 509 U.S. at 593. In a law review article by Professor Mark Green, quoted by the Supreme Court in *Daubert*, falsifiability is further defined as the theory that “knowledge is gained by attempting to disprove or falsify a hypothesis based on empirical investigation. Scientific methodology today is based on generating hypotheses and testing them to see if they can be falsified. Indeed, this methodology is what distinguished science from other fields of human inquiry.” See *id.* at 586 n.4 (citing 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, 645 (1992)).

³⁴ See Gatowski et al., *supra* note 4, at 444-45. This means that judges could not articulate the assumptions underlying *Daubert* (1) that scientific hypotheses are not proved through testing; (2) that testing is aimed at refuting a hypothesis; (3) that hypotheses that are not refutable/falsifiable are by definition not scientific; and (4) that only hypotheses that are not falsified, but could be, provide the basis for reliable scientific theories.

³⁵ The *Daubert* Court cautioned that “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.” *Daubert*, 509 U.S. at 594 (internal citations omitted). When error rates are used to assess the validity of a scientific methodology, they can include false negative errors (when an experimenter misses a real effect), false positive errors (when an experimenter perceives an effect that did not occur), and sampling errors (when an experimenter extrapolates from a small sample to a large population). See KENNETH R. FOSTER & PETER W. HUBER, *JUDGING SCIENCE: SCIENTIFIC KNOWLEDGE AND THE FEDERAL COURTS* 75-76 (1999) (defining false positive and false negative errors and describing how they can result in sampling error).

³⁶ See Gatowski et al., *supra* note 4, at 445.

them, especially with respect to falsifiability and error rate.”³⁷ The researchers also noted that despite specific efforts by interviewers aimed at allowing judges to express their level of comprehension using their own words, “it seems likely that the ambiguity of the [judges’] responses may reflect a genuine lack of understanding of these scientific concepts.”³⁸ It is not surprising that the study found that judges scored much higher in their basic comprehension of the last two *Daubert* criteria: peer review and publication (71%) and the *Frye* criterion of general acceptance (82%).³⁹ However, even more recent scientific data demonstrate that judicial comfort and familiarity with some of the *Daubert* factors may do little to ensure that only valid science enters the courts.

III. FLAWED ASSUMPTIONS: PEER REVIEW MAY NOT ENHANCE THE VALIDITY OF SCIENTIFIC PUBLICATIONS

Of the two remaining *Daubert* criteria, only the requirement that judges assess “peer review and publication” adds any teeth to the old *Frye* inquiry. However, this *Daubert* factor has been dormant for a decade, virtually ignored by courts and commentators.

The *Daubert* Court clearly intended that judges use their *own* evaluation of peer review and publication as a tool for exploring whether substantive flaws in an expert’s methodology have been or could have been exposed.⁴⁰ This reading of *Daubert* makes sense for three reasons. First, it is logically consistent with the Court’s overarching concern that judges play a more active and autonomous role in the evaluation of scientific validity. Second, it fits with the Court’s desire to require judges to assess falsifiability, because judges would be forced to examine peer review and publication in some detail as part of an effort to determine whether a theory is *capable* of being tested. Third, it is consistent with the Court’s concern about error rates because it shares the same goal of exposing methodological flaws. If *Daubert* were read this way, judges would need to look critically at the standards and procedures of any peer review proffered by the expert and evaluate the substance of all published work. It may be no surprise that this is not what has happened. Instead peer review and publication has become a

³⁷ *Id.* at 452.

³⁸ *Id.*

³⁹ *See id.* at 447-48.

⁴⁰ According to the *Daubert* Court, judges must evaluate peer review and publication because “submission to the scrutiny of the scientific community is a component of good science because it increases the likelihood that substantive flaws in the methodology will be detected.”

virtually meaningless *Frye*-type surrogate for real review with the mere fact of peer review, publication, or peer reviewed publication serving as a validity enhancer.⁴¹

This is problematic at both a local and global level. If the fact that a theory or technique has been subjected to peer review and has been published tells us anything useful about its validity, it tells us only something vague and general. The mere fact of peer review and/or publication inevitably tells us nothing about (1) the nature or quality of the peer review process; (2) the effect of peer review on the validity of the methods or conclusions contained in the published work; or (3) whether the validity of the published methods or conclusions is impacted by the manner in which this expert proposes to *use* the theory or technique to make inferences or draw conclusions in this case.

In a series of very recent studies published in the June 5, 2002 issue of JAMA, scientists themselves question whether peer review has any positive effect on the quality of medical articles that appear in even the most prestigious scientific publications.⁴² Here are just a few examples of the conclusions drawn from the many empirical studies assessing the actual value of peer review.

- That an examination of articles before and after they had been subjected to peer review reveals that “true effects [of peer review] have not been determined . . . [and] the term peer review is used to describe a number of different processes . . . [so] it may not always be possible to make a clear distinction between peer review and technical editing.”⁴³
- That a study of methodological problems in published reports of randomized controlled trials (“RCTs”) reveals that “there is considerable evidence that many published reports of RCTs are poor or even wrong, despite their

⁴¹ A recent Lexis search revealed that 92 post-*Kumho* federal appellate court cases mentioned the phrase “peer review and publication.” Eighty-six of these cases were mere recitations of the *Daubert* factors. A recent study assessing the effect of *Daubert* in criminal cases found that appellate court opinions which devoted an average of 1,162 words to a discussion of expert evidence, devoted only an average of 15 words to the peer review and publication criterion. Ironically, cases decided before *Daubert* devoted an average of 21 words to peer review and publication while cases decided after *Daubert* devoted only an average of five words, just one more than repeating the criterion requires. Jennifer L. Groscup et al., *The Effects of Daubert on the Admissibility of Expert Testimony in State and Federal Criminal Cases*, 8 PSYCHOL., PUB. POL’Y & L. 339, 350 (2002).

⁴² See JAMA, *supra* note 26.

⁴³ Tom Jefferson et al., *Measuring the Quality of Editorial Peer Review*, in JAMA, *supra* note 26, at 2787.

clear importance.”⁴⁴

- That a cross-sectional study of 243 original research articles published in a variety of scientific journals reveals that “many published reports are of poor-to-average methodological quality.”⁴⁵
- That “[a]lthough there is considerable evidence that methodological errors are common in articles in medical journals, much published research does not have substantive contribution from a statistician.”⁴⁶
- That a study of the methodological quality of 60 published reports of RCTs reveals that “[t]he validity of . . . results is threatened by the subversion of randomization, resulting in biased allocation to comparison groups, the unequal provision of care apart from the intervention under evaluation, the biased assessment of outcomes, and the inadequate handling of dropouts and losses to follow-up.”⁴⁷

In a summary editorial, Dr. Drummond Rennie explained that these data reflect a sixteen-year long coordinated international effort to research the effects of peer review.⁴⁸ According to Dr. Rennie, despite a concerted effort to study and improve the process, “we find ourselves in the peculiar position of believing still more in the virtues of peer review, a system we know to be time-consuming, complex, expensive, and prone to abuse, while we acknowledge that the scientific evidence for its value is meager.”⁴⁹

The implications of this new empirical evidence are profound. As the world grows more scientifically complex, many judges seem to lack even basic familiarity with the scientific process and scientists themselves have dubious faith in their own quality-control mechanisms. *Daubert*, *Kumho*, and the modifications to Rule 702 reflect a coordinated judicial and legislative effort to construct an analytic tool enabling lay judges (and then in turn lay jurors) to discriminate between good and bad science. This task, which is

⁴⁴ Douglas G. Altman, *Poor-Quality Medical Research: What Can Journals Do?*, in JAMA, *supra* note 26, at 2765.

⁴⁵ Kirby P. Lee et al., *Association of Journal Quality Indicators With Methodological Quality of Clinical Research Articles*, in JAMA *supra* note 26, at 2805.

⁴⁶ Douglas G. Altman et al., *How Statistical Expertise is Used in Medical Research*, in JAMA *supra* note 26, at 2817.

⁴⁷ Karin Huwiler-Müntener et al., *Quality of Reporting of Randomized Trials as a Measure of Methodologic Quality*, in JAMA *supra* note 26, at 2801.

⁴⁸ See Drummond Rennie, Editorial, *Fourth International Congress on Peer Review in Biomedical Publication*, in JAMA, *supra* note 26, at 2759-60.

⁴⁹ *Id.* (quotation omitted).

difficult enough, is exacerbated when judges are hampered by their lack of scientific sophistication and are forced to acknowledge that even published, peer-reviewed studies may contain serious undisclosed methodological flaws.⁵⁰

IV. HOW CAN WE USE THIS INFORMATION?

The question then becomes what impact should these data have on how we frame the admissibility inquiry? Although we should recognize that peer review and publication do not guarantee scientific validity, these quality control mechanisms are not entirely without value. Presumably, scientific journals reject a great deal of bad science. In general, they seek to publish experiments that are described in sufficient detail so that their results may be replicated. Consider the alternative. In a sense the internet is the most pervasive example of unlimited access to information without centralized oversight or control. A quick search of what appear to be medical websites provides “data” to support the following scientific claims: (1) vitamins make people energetic; (2) special diets cure cancer; (3) acupuncture can cure a variety of diseases; and (4) childhood immunizations cause autism.⁵¹ Although imperfect, prepublication peer review maintains minimum standards of honesty and accuracy.⁵² Given the fact that courts cannot abandon or improve the peer review and publication process, the best that judges can do is be more self-conscious about how they understand and use this evidence as part of the validity calculus and recognize that, as a tool, it has unfulfilled potential.

⁵⁰ In another recent empirical study of the effect of *Daubert* on civil cases, researchers concluded that “since *Daubert*, judges have examined the reliability of expert evidence more closely and have found more evidence unreliable as a result. Our analysis, however, does not allow us to conclude whether this increased scrutiny resulted in better outcomes.” Dixon & Gill, *supra* note 17, at 269. “Judges may feel compelled to evaluate reliability and yet not be knowledgeable enough in the relevant field to make accurate determinations.” *Id.* at 301.

⁵¹ For an extensive, regularly updated, and well-organized list of misleading medical and scientific websites, see www.quackwatch.com.

⁵² It is scientifically inaccurate and legally unnecessary to describe scientific theories as valid or invalid. Theories involving science are never wholly accurate or wholly inaccurate explanations of the empirical world. Scientific validity is better understood as a matter of degree rather than in absolute terms. This means that “[v]alidity” in science is not a binary attribute, like pregnancy.” FOSTER & HUBER, *supra* note 37, at 17 (discussing issues of scientific uncertainty and the limited ability of scientists to speak in terms of absolutes). This fits with the legal burden in cases involving scientific evidence, which is that “the proponent of the testimony does not have the burden of proving that it is scientifically correct, but that by a preponderance of the evidence, it is reliable” *Allison v. McGhan Med. Corp.*, 184 F.3d 1300, 1312 (11th Cir. 1999).

V. THE INHERENT FALLACY OF FACT-BASED VALIDITY

Finally, at the global or systemic level, we should recognize that judicial decisions that turn on whether a particular scientific theory or methodology has been reliably *applied* to a given set of facts, are scientifically meaningless. A legal decision that is grounded in fact-specific validity (e.g., a conclusion based on the specific scientific data relied upon by *this* expert to reliably explain *these* facts) is the antithesis of science. Science, in all of its disciplines, is cumulative and based on a continuing aggregation of new data.⁵³ In fact, scientific thought is premised on the assumption that conclusions are valid *only* if they “can be generalized to settings and subjects outside those described in the study.”⁵⁴

Case-specific determinations of the validity of a particular application of a scientific theory or method also undermine our goals by ensuring that little reliable or predictive precedent will be created to assist judges who must decide the admissibility of complex or disputed theories and techniques. We do not want to end up where each court is engaged in a historically isolated fact-based analysis that will not perform the precedential function of controlling the quality of the science and technology that enters our courtrooms.

VI. CONCLUSION

The fact that judges need to evaluate scientific evidence does not make it easy. This is especially true in areas of complex or developing knowledge where the experts themselves are most likely to disagree. The danger is that judges who lack scientific facility and perceive their role as limited to an assessment of the science applied by *this* expert in *this* case may be less rigorous in their evaluation. As the new studies reveal, judges may also be inclined to favor evaluative criteria that are easier for them to comprehend, regardless of their inherent ability to ferret out problems that should detract from the validity of the expert’s opinions and conclusions. Finally, we must acknowledge that case-specific evaluations of science “as applied” run afoul of bedrock scientific principles and widely-accepted

⁵³ See FOSTER & HUBER, *supra* note 37, at 17.

⁵⁴ See *id.* at 146. Professor Allen makes a related observation when he states, “expert testimony cannot advance accurate outcomes locally unless it rests on acceptable epistemological warrant globally. A necessary but not sufficient condition of appropriate testimony ‘locally’ is reliable expertise ‘globally.’” Allen, *supra* note 11, at 6. Similarly, Professor Allen notes “[w]ithout global reliability, one has gibberish. Thus, the logical relationships underlying the Supreme Court’s cases require that both the global and local issues be resolved favorably before an expert should be allowed to testify.” *Id.*

methods of gauging validity. As Justice Rehnquist foretold in his *Daubert* dissent, “I defer to no one in my confidence in federal judges; but I am at a loss to know what is meant when it is said that the scientific status of a theory depends on its ‘falsifiability,’ and I suspect some of them will be, too.”⁵⁵ Apparently, Justice Rehnquist had reason for concern.

⁵⁵ *Daubert*, 509 U.S. at 600 (Rehnquist, C.J., dissenting).