2007

Peer Review and Publication: Lessons for Lawyers

Susan Haack
University of Miami School of Law, shaack@law.miami.edu

Recommended Citation
Susan Haack, *Peer Review and Publication: Lessons for Lawyers*, 36 *Stetson L. Rev.* 789 (2007).

PEER REVIEW AND PUBLICATION: LESSONS FOR LAWYERS*

Susan Haack**

> [A] pertinent consideration [in determining whether a theory or technique is scientific knowledge that will assist the trier of fact] is whether the theory or technique has been subjected to peer review and publication.
>
> —*Daubert v. Merrell Dow Pharmaceuticals, Inc.* (1993)[1]

The phrase "peer review" connotes the evaluation ("review") of scientific or other scholarly work by others presumed to have expertise in the relevant field ("peers"). Specifically, and most to the present purpose, it refers to the evaluation of submitted manuscripts to determine what work is published in professional journals and what books are published by academic presses (in which context it is also called "refereeing," "editorial peer review," or "pre-publication peer review").[2] Occasionally, however, the phrase is used in a much broader sense, to cover the whole long-run history of the scrutiny of a scientist's work within the scientific community, and of others' efforts to build on it,[3] a long-run

---

* This Article was developed from a talk entitled "Scrutinizing Peer Review" given at the National Institute of Justice Conference on Science and the Law held in St. Petersburg, Florida, in September 2005. It appears here by invitation.

** Distinguished Professor in the Humanities, Cooper Senior Scholar in Arts and Sciences, Professor of Philosophy, and Professor of Law, University of Miami.

1. *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 593 (1993) [hereinafter "*Daubert* (1993)"].

2. The phrase also sometimes refers to the evaluation of clinical performance by senior practitioners in a field (in which context it is called "clinical peer review"), to the evaluation of grant proposals to decide what projects are funded (in which context it is called "grant peer review" or "merit review"), and to the evaluation of abstracts, or sometimes submitted papers, to determine what is presented at conferences.

3. "In the broadest sense of the term, *peer review* can be said to have existed ever since people began to identify and communicate what they thought was new knowledge." David A. Kronick, *Peer Review in 18th-Century Scientific Journalism*, 263 JAMA 1321, 1321 (1990).

process of which peer review in the narrower sense is only a small part.

These two conceptions of peer review, the narrow and the broad, both came into play in the arguments over the admissibility of the plaintiffs' expert testimony in *Daubert*.[4] In 1989, granting Merrell Dow's motion for summary judgment on the ground that the Dauberts' proffered causation evidence was inadmissible, the District Court had stressed that "none of the *published* studies show a statistically significant association between the use of Bendectin and birth defects";[5] and affirming this decision in 1991, observing that "no published epidemiological study had demonstrated a statistically significant association between Bendectin and birth defects," and that "the normal peer[-]review process . . . is one of the hallmarks of reliable scientific investigation," Judge Kozinski also took peer-reviewed publication to be a key factor.[6]

But in 1993, when the case came to the Supreme Court, an amicus brief from Chubin et al. criticized the lower courts' reliance on peer-reviewed publication, arguing that "the peer review system is designed to provide a common and convenient starting point for scientific debate, not the final summation of existing scientific knowledge," and that "contrary to the 'generally accepted' myth, publication of an article in a peer review journal is no assurance that the research, data, methodologies, [or] analyses . . . are true, accurate, . . . reliable, or certain or that they represent 'good science.'"[7] And while Justice Blackmun's ruling for the Court included "peer review and publication" as one factor to which courts might look to determine whether expert scientific testimony is "reliable" in the sense required to make it admissible, it did so in a very hedged and cautious way—acknowledging that pre-publication peer review doesn't guarantee "evidentiary reliability" and may hold back well-grounded but innovative work; and that a much better indicator is survival of the long-run

4. Effie J. Chan, Student Author, *The "Brave New World" of* Daubert*: True Peer Review, Editorial Peer Review, and Scientific Validity*, 70 N.Y.U. L. Rev. 100, 113 (1995).

5. *Daubert v. Merrell Dow Pharms., Inc.*, 727 F. Supp. 570, 575 (S.D. Cal. 1989) [hereinafter "*Daubert* (1989)"].

6. *Daubert v. Merrell Dow Pharms., Inc.*, 951 F.2d 1128, 1129, 1131 n. 3 (9th Cir. 1991) [hereinafter "*Daubert* (1991)"].

7. Br. of Daryl E. Chubin et al. as Amici Curiae in Support of Petrs. at 8, 13, *Daubert* (1993), 509 U.S. 579 (typeface altered from original) [hereinafter Amici Br. of Chubin].

scrutiny of the scientific community, i.e., peer review in the broad sense.[8]

Finally, in 1995, ruling on the case on remand from the Supreme Court, and now acknowledging, as Justice Blackmun had, that peer-reviewed publication is no guarantee that testimony is trustworthy, Judge Kozinski argued that nevertheless, the fact "[t]hat the research is accepted for publication in a reputable scientific journal . . . is a significant indication that . . . it meets at least the minimal criteria of good science."[9] So, since "[n]one of the plaintiffs' experts has published his work on Bendectin in a scientific journal," the Court affirmed the lower court's summary judgment once again.[10]

The aim here is to understand how the peer-review process works, how good an indicator it is that scientific testimony is "reliable" in the legally relevant sense, and how courts might best use this *Daubert* factor. So for most of what follows, the focus will be on peer review in the narrow sense—pre-publication peer review. The starting point, in Part I, will be a sketch of the origins of this practice, the ragged process by which it gradually became standard at scientific and medical journals, and the many roles it now plays; the next step, in Part II, will be to articulate the rationale for pre-publication peer review, and the inherent limitations of the system as a quality-control mechanism; and the next, in Part III, will be an exploration of the changes in science, in scientific publication, and in the academy that have put the peer-review system under severe strain, and of some recent instances in which flawed or even fraudulent work has passed peer review.

But in Part IV, an examination of Justice Blackmun's observations about "peer review and publication" in *Daubert*, the broad sense of "peer review" will play a part alongside the narrow. The argument here will be, in brief, that neither Justice Blackmun's observation that peer-reviewed publication is not necessary or sufficient for evidentiary reliability, and that surviving the long-term process of review by the scientific community is a much better indicator of scientific validity, nor his advice to courts—that

---

8. *Daubert* (1993), 509 U.S. at 593.

9. *Daubert v. Merrell Dow Pharms., Inc.*, 43 F.3d 1311, 1318 (9th Cir. 1995) [hereinafter "*Daubert* (1995)"].

10. *Id.* at 1318, 1332.

peer-reviewed publication may be relevant, but is not a dispositive consideration in determining admissibility—is of much practical help.

Subsequently, whether they have excluded testimony in part because it was not based on peer-reviewed publication, or admitted it even though it was not so based, courts seem by and large not to have asked the questions that might throw light on what peer-reviewed publication, or its absence, means in a particular instance. But as we shall see in Part V, a Pennsylvania court's uncommonly common-sense scrutiny of the peer-reviewed Bendectin literature reveals how weak a reed "peer review and publication" can be—and leaves one wondering rather uncomfortably about the way this "*Daubert* factor" got on the legal radar screen in the first place.[11]

I. PRE-PUBLICATION PEER REVIEW: ITS HISTORICAL ROOTS AND PRESENT ROLES

Scientists have always been concerned that their work be acknowledged as theirs and have worried about what Robert Boyle charmingly described as "philosophicall robbery," a.k.a. plagiarism.[12] Even before the *Philosophical Transactions* of the Royal Society of London were inaugurated in 1655, the Society would give its official stamp to a scientist's priority in discovery by recording the date on which it received a letter announcing an experiment or observation.[13] As Henry Oldenburg, the first editor of the *Transactions*, told Boyle, the Society would be "very carefull of registring as well the person and time of any new matter, imparted to ym, as the matter itselfe; whereby the honor of ye invention will be inviolably preserved to all posterity."[14] Gradually the *Transactions* began to indicate which work had and which had not been evaluated by representatives of the Society before publication; and by 1702 the *Journal de Scavans*, founded just before the *Transactions*, had assigned responsibility for screening

11. *Blum v. Merrell Dow Pharms., Inc.*, 764 A.2d 1, 3 (Pa. 2000).

12. Harriet Zuckerman & Robert K. Merton, *Patterns of Evaluation in Science: Institutionalism, Structure, and Functions of the Referee System*, 9 Minerva 66, 70 (1971).

13. *Id.*

14. *Id.* (citing *The Correspondence of Henry Oldenburg* vol. 1, 319 (A. Rupert Hall & Marie Boas Hall eds. & trans., U. Wis. Press 1966)).

submissions in a given area to various members of the editorial board.[15]

In the course of the eighteenth century several other important medical and scientific publications adopted what we would now call "peer review": in 1731, the preface to the first volume of the *Medical Essays and Observations* published by the Royal Society of Edinburgh announced that "[m]emoirs sent by correspondence are distributed according to the subject matter to those members who are most versed in these matters";[16] in 1752, the Royal Society set up a committee authorized to call on "any other members of the Society who are knowing and well skilled in that particular branch of Science that shall happen to be the subject matter" of an article submitted to the *Transactions*;[17] in 1782, the regulations of the Académie Royale de Médecine stated that "[n]othing will be printed in the *Histoire*, or in the *Receuil des memoires* of the Society . . . which assemblies especially called for this purpose have not decided by a majority vote to publish";[18] and in 1785, the Literary and Philosophical Society of Manchester set up a reviewing committee to select papers "with as much impartiality, and as strict attention to their comparative merits" as possible.[19]

According to historian John Burnham, the spread and evolution of the practice of pre-publication peer review through the nineteenth century and the early decades of the twentieth was neither systematic nor orderly.[20] Some of the earliest medical journals of the nineteenth century were, as Burnham puts it, "personal vehicle[s]" for editors like Thomas Wakely, founder of *The Lancet*, or Henry Maunsell, a founder of the *Dublin Medical Press*, who subsequently also became owner of the *Dublin Eve-*

---

15. Stephen Lock, *A Difficult Balance: Editorial Peer Review in Medicine* 2 (Nuffield Provincial Hospitals Trust 1985).

16. Kronick, *supra* n. 3, at 1321 (citing *I Essais et observations de médecine de la Société d'Edinbourg* vol. 1, preface (1740)).

17. *Id.* (citing J.M. Ziman, *Information, Communication, Knowledge*, 224 Nature 318, 318 (1969)).

18. *Id.* at 1321–1322 (citing 13 *Histoire de l'Académie Royale de Médecine* 19–21 (1782)).

19. *Id.* at 1322 (citing 1 *Memoirs of the Literary and Philosophical Society of Manchester* preface (1785)).

20. John C. Burnham, *The Evolution of Editorial Peer Review*, 263 JAMA 1323, 1327–1328 (1990).

ning Mail.[21] Somewhat closer to present-day scientific and medical journals were the official publications of European (especially German) research institutes; these more specialized fora relied on the expert judgment of the editor or the colleagues who made up his editorial staff, but were essentially outlets for the work of members of the institute.[22]

But in the early days of both scientific and medical publishing an editor's problem was more likely to be finding enough material to fill his pages than deciding which of too many articles to publish. In 1876, a commentator observed that "... the demand for brief papers and reports of single cases, exceeds the supply.[23] The weekly and monthly periodicals are omnivorous and insatiable in their requests for contributions";[24] even in 1921 the editor of *The Journal of Neurology and Psychopathology* was complaining to a correspondent about the difficulty of getting enough material for his journal.[25] It was only after World War II that peer review as we now know it became common practice in medical and scientific journals;[26] for by this time a significant shift in the number of papers offered meant that editors were looking, not for material to fill their pages, but for a way to select which papers they would publish.[27]

By now, pre-publication peer review is routine at medical and scientific journals,[28] and standard procedure, too, in scholarly publication and in other areas, including the humanities (though not the law reviews).[29] It has, in consequence, also become a very

---

21. *Id.* at 1324.

22. *Id.*

23. *Id.* at 1325.

24. *Id.* (citing John Shaw Billings, *Literature and Institutions*, 72 Am. J. Med. Sci. 439, 460 (1876)).

25. *Id.* (citing Ltr. from C. Stanford Read to Smith Ely Jelliffe (Feb. 3, 1921), in *Papers of Smith Ely Jelliffe, 1866–1940* (on file with Library of Congress, Washington, D.C., Box 16)).

26. Marjorie Sun, *Peer Review Comes under Peer Review*, 244 Sci. 910, 910 (1989). James McKeen Cattell, who edited *Science* from 1894 until his death in 1945, reportedly relied heavily on his son (who had a degree in physiology from Harvard) to help screen submissions; but when the American Association for the Advancement of Science took over the journal in 1945, a system of peer reviewing was instituted. *Id.*

27. Burnham, *supra* n. 20, at 1326–1327.

28. By 1985 at least three-quarters of major scientific journals in the West relied on peer review. Lock, *supra* n. 15, at 3. In 1980, the 100 Soviet medical journals also used peer review. *Id.*

29. At law reviews it is usually student editors, not faculty, who decide what papers

important factor in the economics of medical, scientific, and other academic publishing; the prestige of the big scientific and medical publishing houses and of the academic presses, and hence the high prices they can command for their publications,[30] derives in part from these publications' being perceived as somehow "certified" by peer review.

Moreover, peer review is now deeply entrenched in the tenure and promotion systems of the universities, which may require peer-reviewed publications or look less favorably on publications that are not peer-reviewed, and may count a faculty member's acting as referee for scholarly journals or presses as part of his or her "service."[31] In fact, universities often use pre-publication peer review as a proxy—it is tempting to say, as a lazy substitute—for substantive assessment of the quality of a person's work. As an unusually candid editorial in *Nature* complained, "universities . . . have slipped into the sloppy habit of substituting for their own judgement of their own achievements the judgement of external assessors as delivered by the appropriate sub-net of the peer-review system."[32]

As Percy Bridgman once observed, while "[a] dog is content to turn around three times before lying down," a human being would have to think up some reason why this is the *best* way to lie down; "[t]here is not a single human institution which has not originated in hit or miss fashion, but, nevertheless, every one of these institutions is justified by some rationalizing argument as the best possible."[33] So it is no surprise that, as pre-publication peer

---

are accepted. *See* Richard A. Posner, *Against the Law Reviews*, 2004 Leg. Affairs 57, 57 (Nov./Dec. 2004) (acknowledging that student editors ultimately decide which articles to publish and arguing that the law review publication process is, for this reason, less rigorously controlled than publication in other academic fields); *but see* Amici Br. of Chubin, *supra* n. 7, at 8 n. 8 (pointing out that law reviews are in some respects more rigorous, since student editors, who check every citation and footnote, spend far more time on papers than peer reviewers for scientific journals can do).

30. In October 2003 scientists at the University of California, San Francisco staged a protest over Elsevier's $91,000 bill for six biology journals; eventually the university negotiated "a 25% price reduction to $7.7 million a year for 1,200 Elsevier periodicals." Bernard Wysocki Jr., *Peer Pressure: Scholarly Journals' Premier Status Is Diluted by Web*, Wall St. J. A1, A8 (May 23, 2005).

31. Editorial, *Is Science Really a Pack of Lies?* 303 Nature 361, 361 (1983).

32. *Id.*

33. Percy Bridgman, *The Struggle for Intellectual Integrity*, in *Reflections of a Physicist* 361, 368 (2d ed., Phil. Lib 1955).

review has spread and become entrenched in academic publishing and in the academy itself, some are tempted to exaggerate its virtues—to think of the system, not just as a rough-and-ready preliminary filter, but as a strong indication of quality. In 1968, John Ziman described the referee as "the lynchpin [*sic*] about which the whole business of science is pivoted"; and more recently, life scientist Paul Gross writes that he sees "peer-reviewed" as—speaking "loosely, but not incorrectly"—a kind of "antonym" for "biased."[34]

But even if the pre-publication peer-review system worked perfectly, it would be inherently limited in what it could do to ensure quality—of which, in any case, "reliability" in the legally relevant sense is only one dimension; what's more, there is good reason to fear that, because of changes in the scale and culture of the sciences since the system became standard, the system now works very imperfectly indeed.

II. PRE-PUBLICATION PEER REVIEW: ITS RATIONALE AND INHERENT LIMITATIONS

In 1946, just as the practice was becoming standard procedure at scientific journals, Michael Polanyi gave the classic statement of an epistemological rationale for pre-publication peer review. Some system for rationing limited publication opportunities is essential, he argued, for the scientific enterprise depends on effective evidence-sharing and mutual scrutiny, and without such a system scientists will be obliged to waste their time sifting through the work of cranks and incompetents looking for worthwhile stuff as follows:

> Suppose . . . that no limitations of value were imposed on the publication of scientific contributions in journals. The selection—which is indispensable in view of the limited space—would then have to be done by some neutral method—say drawing lots. Immediately the journals would be flooded with rubbish and valuable work would be crowded out. . . . Cranks are always abounding who will send in spates of nonsense. Immature, confused, fantastic, or else plodding,

34. E-mail from Paul R. Gross, Prof. of Life Scis., U. of Va., to Susan Haack, *Peer Review* (July 11, 2005, 3:46 p.m. EDT) (copy on file with the Author).

> pedestrian, irrelevant material would be pouring in. Swindlers and bunglers combining all variants of deception and self-deception would seek publicity. Buried among so much that is specious or slipshod, the few remaining valuable publications could hardly have a chance of being recognized.[35]

Rationing by pre-publication peer review, Polanyi continued, is a way to ensure that what is published at least meets minimal standards of professional competence as follows:

> No proposed contribution to science has a chance of becoming generally known unless it is published in print; and its chances of recognition are very poor unless it is published in one of the leading scientific journals. The referees and editors of these journals are responsible for excluding all matter which they consider unsound or irrelevant. They are charged with guarding a minimum standard for all published scientific literature.[36]

The key phrases for our purposes are "unsound or irrelevant" and "guarding a minimum standard."[37]

"Unsound" and "minimum standard" make the point that prepublication peer review cannot be expected to guarantee truth, sound methodology, rigorous statistics, etc. From the very beginning, scientific editors have stressed that they and their reviewers have no choice but to rely on the integrity of authors. In 1665, Denis de Sallo, the first editor of the *Journal des Scavans*, wrote in the first issue that "we aim to report the ideas of others without guaranteeing them";[38] the Edinburgh Society's 1731 statement of its refereeing policy concludes with the observation that "[r]esponsibility concerning the truth of facts, the soundness of reasoning, in the accuracy of calculations is wholly disclaimed: and must rest alone, on the knowledge, judgement, or ability of the authors who have respectfully furnished such communications."[39]

---

35. Michael Polanyi, *Science, Faith and Society* 35–36 (Oxford U. Press 1946).
36. *Id.* at 33.
37. *Id.*
38. Drummond Rennie, *Editorial Peer Review: Its Development and Rationale*, in *Peer Review in Health Sciences* 1, 2 (Fiona Godlee & Tom Jefferson eds., 2d ed., BMJ Publg. Group 2003).
39. Kronick, *supra* n. 3, at 1322.

And Polanyi's "irrelevant" reminds us that editors and peer-reviewers are not concerned *only* with truth, methodological soundness, and such; they also care, reasonably enough, about the interest of the work, the readability of the article, and its suitability for this particular journal.[40] As the former editor of the *Journal of the National Cancer Institute* puts it, writing of "[r]eliability . . . and other inappropriate goals in peer review," "editorial decisions can, do, and should make use of criteria . . . [such as] originality, the suitability of the topic for a given journal, . . . the need for a balance of topics in journals with broad coverage, [and] the importance of findings to readers . . . ."[41]

Polanyi was clear that what gives scientific results some authority is not peer-reviewed publication as such, but the following that happens after work is published:

> On its publication a paper is laid open to scrutiny by all scientists who will proceed to form, and possibly also to express, an opinion on its value. They may doubt or altogether reject its claims, while its author will probably defend them. After a time a more or less settled opinion will prevail. The third stage of public scrutiny through which a contribution to science must pass in order to become generally known and established is its incorporation in text-books or at least standard books of reference.[42]

Moreover, he acknowledged that the peer-review system will succeed even in the modest task of "guarding a minimum standard" only on certain following assumptions:

> If each scientist set to work every morning with the intention of doing the best bit of safe charlatanry which would just help him into a good post, there would soon exist no effective standards by which such deception could be detected. . . . Only if scientists remain loyal to scientific ideals

40. *See* Polanyi, *supra* n. 35, at 33 (suggesting that referees and editors try to ensure that all published literature meets minimum standards of quality and is relevant to the journal).

41. John C. Bailar, *Reliability, Fairness, Objectivity and Other Inappropriate Goals in Peer Review*, 14 Behavioral & Brain Scis. 137, 138 (1991).

42. Polanyi, *supra* n. 35, at 33–34.

> rather than try to achieve success with their fellow scientists can they form a community which will uphold these ideals.[43]

Obviously (though Polanyi doesn't say so explicitly), the effectiveness of the system depends not only on the integrity of authors, but also on the integrity of reviewers, editors, and publishers. And the problem is not only that it will fail if every scientist sets to work to do "the best bit of safe charlatanry" he can get away with; it is also that peer review will function less effectively the heavier the burdens on reviewers and editors, the greater the pressures on journals, and the greater the temptations for scientists to cut corners, or to fudge, trim, or even fake results.

III. PRE-PUBLICATION PEER REVIEW: RECENT STRESSES AND STRAINS, FLAWS AND FAILURES

Even in ideal circumstances, reviewers are better placed to judge the readability of a paper or the interest of its topic or results than its truth or accuracy, and may in good faith reject important work that is too innovative to seem plausible; so perhaps it is not surprising that by 1994 historian of science Horace Freeland Judson, describing the "structural transformations . . . taking place in the sciences," included "declining standards and the growing, built-in tendency toward corruption of the peer-review and refereeing processes" on his list.[44] For today there are many pressures putting the peer-review system under severe strain: the explosion of scientific and medical publications; the increasing financial influence of large drug companies on the medical journals; the pressures on young scientists to get grants and to publish; the temptations to celebrity-seeking; the burgeoning expert-witness business; and so on.[45]

There are variations among the scientific and medical journals, but the peer-review refereeing process usually works roughly like this: an editor carries out what Lock describes as "triage": "classifying articles into self-evident masterpieces, obvi-

---

43. *Id.* at 40.

44. Horace Freeland Judson, *Structural Transformations of the Sciences and the End of Peer Review*, 272 JAMA 92, 92 (1994).

45. Susan Haack, *Defending Science—Within Reason: Between Scientism and Cynicism* 27–29, 107–109 (Prometheus Bks. 2003).

ous rubbish, and the remainder . . . needing careful consideration";[46] for this third group—the large majority—the editor then chooses one or two (seldom more) reviewers to look at each paper chosen, generally informing reviewers of authors' names, but not vice versa;[47] reviewers are usually given a checklist against which to check for various aspects of style and presentation and certain kinds of obvious error;[48] the reviewers are given a time limit, often of no more than two weeks, to respond with their assessment and recommendation;[49] and they spend an average of around 2.4 hours evaluating a manuscript—which usually involves, not simply giving a "yes or no" verdict, but making suggestions as to how the paper might be improved.[50] Many journals don't check the statistical calculations in accepted papers;[51] and reviewers are in

---

46. Lock, *supra* n. 15, at 6.

47. Some journals are moving towards "open" review in which authors also know reviewers' names. *See* Richard Smith, *Peer Review: Reform or Revolution?* 315 British Med. J. 759, 760 (1997) (arguing that open review is the most ethical form because it places authors and reviewers in equal positions and allows for increased accountability). By contrast, in philosophy journals, and so far as I know in humanities journals generally, *both* reviewers' *and* authors' names are normally "blinded." (I routinely decline to referee a paper if the author is known to me.)

48. Water Resources Research Inst. of U.N.C., *Report Guidelines for Authors—Editorial Checklist*, http://www.ncsu.edu/wrri/reports/guidelines/edchecklist.html (accessed Oct. 1, 2006).

49. In philosophy journals, and so far as I know in humanities journals generally, the time allowed is much longer.

50. *See e.g.* Stephen Lock & Jane Smith, *What Do Peer Reviewers Do?* 263 JAMA 1341, 1342 (1990) (indicating that study results show that reviewers spend less than 2 hours reviewing a manuscript); Alfred Yankauer, *Who Are the Peer Reviewers and How Much Do They Review?* 263 JAMA 1338, 1339 (1990) (reporting that for 12 issues of *American Journal of Public Health* the average review time was 2.4 hours that resulted in 3360 hours of uncompensated time).

51. *See* Martin J. Gardner & Jane Bond, *An Exploratory Study of Statistical Assessment of Papers Published in* The British Medical Journal, 263 JAMA 1355, 1355 (1990) (quoting statistics from a study on accuracy of papers submitted to *The British Medical Journal*; only 11% of submitted papers were found to be statistically accurate, and only 84% of published papers were accurate); Ann C. Weller, *Editorial Peer Review in US Medical Journals*, 263 JAMA 1344, 1345 (1990) (reporting that most journals do not make any independent check of authors' statistical calculations); *see also* Dianne Bryant et al., *How Many Patients? How Many Limbs? Analysis of Patients or Limbs in the Orthopedic Literature: A Systematic Review*, 88 J. Bone & Joint Surgery 41, 41 (2006) (concluding that 42% of clinical studies in highly-rated orthopedic journals are biased by the inclusion of multiple observations of different limbs of single individuals); Emili García-Berthou & Carles Alcaraz, *Incongruence between Test Statistics and* P *Values in Medical Papers*, 4 BMC Medical Research Methodology 13, "Results and Discussion" (2004) (finding that "11.6% (21 of 181) and 11.1% (7 of 63) of the statistical results published in *Nature* and *BMJ* respectively during 2001 were incongruent" and noting that "[a]t least one such error ap-

no position to repeat authors' experiments or studies, which will ordinarily have taken a good deal of time and/or money. Acceptance rates vary widely from field to field; where the rate is low, most of the papers initially submitted to but rejected by one or more of the most desirable journals eventually appear in some lower-ranked publication, and a paper "may have been rejected by ten or twenty journals before it is finally accepted."[52] Textbook chapters are usually invited, not peer-reviewed. Nor are all the articles in "peer-review" journals peer-reviewed; some are invited, and some appear by editorial privilege; and sometimes the authors have been asked—as I have been asked myself—to nominate their own reviewers.

As the scale of the operation increases, with more and more papers submitted to more and more journals, the quality of reviewers and the time and attention they can give to their task is likely to decline. As the career pressures on scientists intensify, the temptation grows for reviewers to recommend acceptance of work they perceive as likely to advance their careers, to recommend rejection of work they perceive as a professional threat, or to plagiarize ideas from work they are asked to review.[53] And as pressures on the journals and their staff increase, the hope of prestige and profit causes further distortions: some journals suspend the peer-review process when they publish symposia sponsored by pharmaceutical companies (for which the journal may charge the company a significant fee); some reap large sums from the sale of large numbers of reprints to the companies con-

---

peared in 38% (12 of 32) and 25% (3 of 12) of the papers of *Nature* and *BMJ* respectively, indicating that they are widespread and not concentrated in a few papers"); Julie A. Neville et al., *Errors in the* Archives of Dermatology *and the* Journal of the American Academy of Dermatology *from January through December 2003*, 142 Archives Dermatology 737, 738 (2006) (reporting that from January through December 2003, the *Archives of Dermatology* and the *Journal of the American Academy of Dermatology* published 364 studies where "59 (38.1%) of 155 [that used statistical analysis] contained errors or omissions in statistical methods or the presentation of the results"); A. Vail & E. Gardener, *Common Statistical Errors in the Design and Analysis of Subfertility Trials*, 18 Human Reprod. 1000, 1000 (2003) (reporting that of thirty-nine trials studied, "[s]ix trials were fatally flawed by design" and "[o]nly five trials reported live birth rates sufficiently to allow valid meta-analysis").

52. Amici Br. of Chubin, *supra* n. 7, at 16.

53. D.H. Osmond, *Malice's Wonderland: Research Funding and Peer Review*, 14 J. Neurobiology 95, 105 (1983).

cerned;[54] some put pressure on authors to cite other papers in the same journal, thus raising its "impact factor" and boosting library orders;[55] and so on.

Editors themselves have begun to express concern. Richard Smith, editor of *The Lancet*, writes that peer review is "expensive, slow, prone to bias, open to abuse, possibly anti-innovatory, and unable to detect fraud."[56] Drummond Rennie, associate editor of the *Journal of the American Medical Association* (JAMA), is even more outspoken: "[t]here seems to be no study too fragmented, no hypothesis too trivial, no literature citation too biased or too egotistical, no design too warped, no methodology too bungled, . . . no argument too circular, no conclusion too trifling or too unjustified, and no grammar or syntax too offensive for a paper to end up in print."[57]

* * *

According to a study reported in JAMA in 2004, a survey of 122 published articles found that fifty percent of efficacy and sixty-five percent of harm outcomes were incompletely reported.[58] According to a study reported in *Nature* in 2005, more than ten percent of 3,247 scientists polled admitted withholding details of methodology or results from papers or proposals; more than fifteen percent admitted dropping observations or data points; and more than twenty-seven percent admitted keeping inadequate records of research projects.[59] According to a study reported in JAMA the same year, of forty-five highly cited studies published in prestigious journals and claiming effective medical interventions, fourteen were later contradicted in whole or part by other

---

54. "Two editors reported that their journals charged $400 to $1,500 per page to publish symposiums, and another reported charging a flat fee of $100,000. The journals charged an average of $15 per reprint, and reprint requests for symposiums [averaged] 25,000." Lisa A. Bero et al., *The Publication of Sponsored Symposiums in Medical Journals*, 327 New Eng. J. Med. 1135, 1136–1137 (1992).

55. Sharon Begley, *Science Journals Artfully Try to Boost Their Rankings*, Wall St. J. B1, B8 (June 5, 2006).

56. Smith, *supra* n. 47, at 759.

57. Drummond Rennie, *Guarding the Guardians: A Conference on Editorial Peer Review*, 256 JAMA 2391, 2391 (1986).

58. An-Wen Chan et al., *Empirical Evidence for Selective Reporting of Outcomes in Randomized Trials*, 291 JAMA 2457, 2457 (2004).

59. Brian C. Martinson et al., *Scientists Behaving Badly*, 435 Nature 737, 737 (2005).

studies.[60] And according to an article published in the Rockefeller University's *Scientist* magazine in spring 2006, over the three years in which the *Journal of Cell Biology* has been examining every image in every paper accepted, checking for alterations made in Adobe Photoshop, 14 of 1,400 articles were rejected after fraudulent image alteration was detected.[61]

Moreover, other studies suggest that even after serious scientific misconduct or outright fraud has been discovered, the process of cleaning up the scientific literature so that such work is retracted and others' innocent citations to it corrected is at best patchy and uneven.[62] For example, a year after the Office of Research Integrity informed ten journals that papers co-authored by Dr. Eric Poehlman they had published were fraudulent, only eight had retracted; and even after the *Annals of Internal Medicine* had retracted one of these papers, other authors continued to cite it.[63]

* * *

In fact, there are so many recent reports of failures of the peer-review system that the difficulty is to select the most instructive. Should it be the notorious case of Dr. Hwang Woo Suk, the researcher whose apparently stunning work on cloning, published in *Science* and *Nature*, turned out to rest on fabricated data?[64] Or should it be that extraordinary article in the *Journal of Reproductive Medicine*, claiming to have shown that intercessory

---

60. John P.A. Ionnadis, *Contradicted and Initially Stronger Effects in Highly Cited Clinical Research*, 294 JAMA 218, 220 (2005).

61. Lauren Gravitz, *Biology's Image Problem*, Rockefeller U. Scientist 1, 10 (Spring 2006).

62. Paul J. Friedman, *Correcting the Literature following Fraudulent Publication*, 263 JAMA 1416, 1417 (1990); Mark P. Pfeiffer & Gwendolyn L. Snodgrass, *The Continued Use of Retracted, Invalid Scientific Literature*, 263 JAMA 1420, Abstract (1990).

63. Jennifer Couzin & Katherine Unger, *Cleaning up the Paper Trail*, 312 Sci. 38, 39 (2006); Harold C. Sox & Drummond Rennie, *Research Misconduct, Retraction, and Cleansing Medical Literature: Lessons from the Poehlman Case*, 144 Annals Internal Med. 609, 609 (2006) (noting that in 1989, in order to evaluate allegations of scientific fraud, Congress created the Office of Scientific Integrity, later renamed the Office of Research Integrity). The article that was retracted was Eric T. Poehlman et al., *Changes in Energy Balance and Body Composition at Menopause: A Controlled Longitudinal Study*, 123 Annals Internal Med. 673 (1995). Sox & Rennie, *supra* n. 63, at 609.

64. Nicholas Wade & Choe Sang-Hun, *Human Cloning Was All Faked, Koreans Report*, N.Y. Times A1 (Jan. 10, 2006) (quoting Dr. Benjamin Lewin, former editor of *Cell*, commenting that *Science* should have been more careful and certainly shouldn't have published a paper with "several identical photos").

prayer by strangers of another faith in another country doubled the success rate of attempted *in vitro* fertilizations—the supposed lead author of which learned of the study only six to twelve months after it was completed, and another, a law school graduate with no medical degree or scientific training, subsequently pled guilty to (unrelated) charges of business fraud?[65] Or maybe the papers by Jon Sudbø in *The Lancet* and the *New England Journal of Medicine* (NEJM), claiming to have shown that non-steroidal anti-inflammatory drugs reduced the risk of oral cancer, all of which turned out to have been based on fabricated data?[66] Or should it be something lower-key, such as the article in the NEJM, cited in litigation against Metabolife, in which the information in a table of eleven patients listing adverse effects and pre-existing conditions is contradicted by the text on the very same page?[67]

But no: the extraordinary saga of the report of Merck's large-scale clinical trial of Vioxx, the VIGOR study—on the basis of which, in May 1999, the FDA approved the drug for sale—stands

65. Benedict Carey, *Researcher Pulls His Name from Paper on Prayer and Fertility*, N.Y. Times A15 (Dec. 4, 2004); Bruce Flamm, *The Columbia University 'Miracle' Study: Flawed and Fraud*, 28 Skeptical Inquirer 25, 27–28 (Sept./Oct. 2004). Thanks to Elizabeth Balbin for this example. The article concerned was Kwang W. Cha et al., *Does Prayer Influence the Success of* In Vitro *Fertilization-Embryo Transfer? Report of a Masked, Randomized Trial*, 46 J. Reprod. Med. 781, 782 (2001). Flamm, *supra* n. 65, at 27–28.

66. The database of 908 participants in the *The Lancet* study, reportedly, was simply made up; 250 of the fictional persons involved supposedly had the same birth date! Forbes, *Many Researches Break the Rules: Study*, http://www.forbes.com/forbeslife/health/feeds/hscout/2006/04/13/hscout532110.html (Apr. 13, 2006). *See also* Richard Horton, *Expression of Concern: Non-Steroidal Anti-Inflammatory Drugs and the Risk of Oral Cancer*, 367 Lancet 196, 196 (2006) (expressing concern over verbal admission by Sudbø that he fabricated data for the study previously published in *The Lancet* and acknowledging possible misconduct in two of his research papers published in the NEJM); Richard Horton, *Retraction—Non-Steroidal Anti-Inflammatory Drugs and the Risk of Oral Cancer: A Nested Case-Control Study*, 367 Lancet 382, 382 (2006) (retracting a Sudbø article from a previous issue of *The Lancet* based on confirmation that data was fabricated). The article retracted from *The Lancet* was John Sudbø et al., *Non-Steroidal Anti-Inflammatory Drugs and the Risk of Oral Cancer: A Nested Case-Control Study*, 366 Lancet 1359 (2005). *Id.*

67. Christine Haller & Neal L. Benowitz, *Adverse Cardiovascular and Central Nervous System Events Associated with Dietary Supplements Containing Ephedra Alkaloids*, 343 New Eng. J. Med. 1833, 1836 (2000). Table four on page 1836 lists patient number seven as having no pre-existing conditions or concurrent risks, yet the text on the same page indicates that an autopsy of this patient "showed mild cardiomegaly with four-chamber dilatation and coronary artery disease, with narrowing of 50 to 75 percent in four vessels." Thanks to Dr. Robert Myerburg for this example.

out as an object-lesson in what can go wrong.[68] After FDA approval, the report of the study—concluding that Vioxx carried a lower risk of adverse gastrointestinal effects than older pain-relievers, and that for most patients its risk of adverse cardiovascular effects was not significant—was submitted to the NEJM, where it appeared in November 2000.[69] In 2002, however, Merck was obliged to add a warning about cardiovascular risks to the package insert. And in September 2004—after the data safety monitoring board halted another major clinical trial, the APPROVe study (designed to show that Vioxx lowered the risk of colon polyps), when it emerged that patients given twenty-five milligrams of Vioxx for more than eighteen months had a fourfold greater incidence of serious thromboembolic events—Merck withdrew the drug from the market.[70]

In December 2005, in the midst of a gathering storm of litigation by patients claiming they had been injured by the drug, the NEJM issued an "Expression of Concern" acknowledging that three heart attacks among patients taking Vioxx had been omitted from the report of the VIGOR study it had published in 2000.[71] These adverse events had been included in the data on the FDA website since February 2001; and two of the three authors had known of them well in advance of the publication of the paper.[72] Their inclusion raised the rate of heart attacks among those taking Vioxx from the 0.4% claimed in the paper to 0.5% (compared to 0.1% among patients taking naproxen) and moreover contradicted the claim in the paper that only those already at risk showed an increase in heart attacks after taking Vioxx.[73] Merck claimed that the additional heart attacks occurred after the cut-off date for the study; but the editor of the journal, Dr. Jeffrey

68. Sen. Comm. on Fin., *FDA, Merck, and Vioxx: Putting Patient Safety First?* 108th Cong. (Nov. 18, 2004) (testimony of Dr. Sandra Kweder, Dep. Dir., Off. of New Drugs, Ctr. for Drug Evaluation and Research, FDA).

69. Claire Bombadier et al., *Comparison of Upper Gastrointestinal Toxicity of Rofecoxib and Naproxen in Patients with Rheumatoid Arthritis*, 343 New Eng. J. Med. 1520 (2000).

70. Simon R.J. Maxwell & David J. Webb, *Cox-2 Selective Inhibitors—Important Lessons Learned*, 365 Lancet 449, 449 (2005).

71. David Armstrong, *Bitter Pill: How the* New England Journal of Medicine *Missed Warning Signs in Vioxx*—Medical Weekly *Waited Years to Report Flaws in Article that Praised Pain Drug—Merck Seen as 'Punching Bag'*, Wall St. J. A1, A10 (May 15, 2006).

72. *Id.*

73. *Id.*

Drazen, told reporters that the design of the study, which continued to track gastrointestinal effects after it stopped tracking cardiovascular effects, had been misleading.[74]

But the problem here wasn't only with the authors; nor was it only that the journal's reviewers didn't have the raw data or that they failed to notice the oddity in the study design. We now know that in June 2001 the editors of the NEJM had received a letter from pharmacist Jennifer Hrachovec asking that the article be corrected in light of the information on the FDA website, but had declined to publish it on the grounds that the journal "can't be in the business of policing every bit of data we put out";[75] that when deposed by the parties in federal litigation in Texas in November 2005, executive editor Dr. Gregory Curfman acknowledged that neither the reviewers nor the editors had questioned Merck's theory that the higher rate of cardiovascular events among Vioxx patients was attributable to a cardio-protective effect of naproxen, even though an FDA official had noted that it "is not supported by any . . . controlled trials";[76] that the journal had sold 929,000 copies of reprints of the article, most of them to Merck, for revenue estimated to be between $697,000 and $836,000; and that the "Expression of Concern" about the study had been published on the urgent last-minute advice of public-relations specialist Edward Cafasso that testimony to be presented the next day in the Vioxx case in which Dr. Curfman had been deposed made it essential for the journal to post something right away, to "drive the media away from NEJM and toward the authors, Merck and plaintiff attorneys."[77] As Richard Smith, former editor of the *British Medical Journal*, observed, the conduct of the NEJM in the dispute over the VIGOR trial "raised doubts about the journal's

74. *Id.*

75. *Id.* (quoting Dr. Jeffrey Drazen, a top editor for the *New England Journal of Medicine*, during a radio appearance in August 2001).

76. *Id.*

77. *Id.* The "Expression of Concern" was published on-line on December 8, 2005, the day the jury began deliberations in the third Vioxx trial. Diedtra Henderson, *Journal Says Vioxx Woes Suppressed; Merck Blamed; Correction Sought*, Boston Globe A1 (Dec. 9, 2005). According to Henderson, in December 2005, Dr. Curfman said that the NEJM had "learned of the new information [i.e., the three omitted heart attacks] about two weeks [earlier]"; according to Armstrong, *supra* n. 71, at A1, however, the journal had known about them at least since Ms. Hrachovec's letter in June 2001.

integrity"; "[t]he journal failed its readers [and] damaged its reputation."[78]

And just as you thought it could hardly get worse, in July 2006 the NEJM published a correction to the report it had earlier published of the APPROVe study: key results claimed in the report had not been arrived at by the statistical method the authors said they used; moreover, using the method the authors had said they were using, but had not in fact used, the results undermined the claim in the report that cardiovascular risks increased only after eighteen months.[79]

Not long before, Lawrence Altman had written in the *New York Times* that "[r]ecent disclosures of fraudulent or flawed studies in medical and scientific journals have called [the peer-review system] into question as never before . . . ";[80] it is hard to disagree.

* * *

For obvious reasons they are harder to track, and for obvious reasons they are often not known until long after the event; but it is pretty clear that there are also many instances in which important and innovative work has been rejected by peer-reviewers. Lock tells the story of Edward Jenner's report of his smallpox vaccination, which was rejected by the *Transactions* of the Royal Society in 1796, after Sir Joseph Banks had looked it over and reported that he "wanted faith" in its conclusion.[81] Charles McCutchen, lamenting the way "[r]eviewing weeds out good manuscripts as well as poor ones," lists "Frederick Lanchester's 1894 circulation theory of how wings lift, Chandra Bose's photon statistics in 1924, Enrico Fermi's theory of beta decay in 1933, Herman Almquist's discovery of vitamin K_2 in 1935, Hans Krebs'

78. Med. News Today, New England Journal of Medicine *Damaged by Its Conduct over Vioxx, Says Former Editor of* British Medical Journal, www.medicalnewstoday.com/medicalnews.php?newsid=46831 (July 9, 2006).

79. Heather Won Tesoriero, *Vioxx Study Correction May Add Pressure to Merck's Defense*, Wall St. J. A2 (June 27, 2007). The original article was Robert S. Bresalier et al., *Cardiovascular Events Associated with Rofecoxib in a Colorectal Adenoma Chemoprevention Trial*, 352 New Eng. J. Med. 1092 (2005). *Correction*, 355 New Eng. J. Med. 221, 221 (2006). Ironically enough, at page 1131, the same issue of the journal includes a short paper by Jeffrey M. Drazen entitled *COX-2 Inhibitors—A Lesson in Unexpected Problems.*

80. Lawrence K. Altman, *For Science's Gatekeepers, a Credibility Gap*, N.Y. Times F1 (May 2, 2006).

81. Lock, *supra* n. 15, at 2.

citric-acid cycle in 1937, and Raymond Lindeman's trophic-dynamic concept in ecology in 1941"; all were "turned down at least once."[82] David Horrobin adds that Krebs' paper, "possibly the most important single article in modern biochemistry, . . . eventually led to a Nobel prize";[83] and lists many other examples, including: a "seminal paper[ ] in immunology" by Glick et al. on the identification of B lymphocytes, which "was rejected by leading general and specialist journals and eventually appeared in *Poultry Science* because of the species on which the work was done";[84] and a paper by New Zealand farmer Gladys Reid suggesting that facial eczema in sheep might be caused by a marginal zinc deficiency, which was rejected by the journals in the field until Horrobin published it in *Medical Hypotheses*—after which her work was confirmed, the disease was eliminated, and Ms. Reid was awarded a decoration for services to New Zealand agriculture.[85]

By now it should hardly need saying: *the fact that work has passed pre-publication peer review is no guarantee that it is not flawed or even fraudulent; and the fact that work has been rejected by reviewers is no guarantee that it is not an important advance.*

IV. LESSONS FOR LAWYERS

"Enough already!" you may be thinking. To be sure, Judge Kozinski's confidence that "the normal peer[-]review process . . . is one of the hallmarks of reliable scientific investigation" was over-optimistic;[86] but didn't Justice Blackmun clear all this up, more than a decade ago, in his majority opinion in *Daubert*?[87]

Well, evidently Justice Blackmun paid attention to the brief from amici Chubin et al.,[88] for he acknowledged the following:

---

82. Charles W. McCutchen, Peer Review: *Treacherous Servant, Disastrous Master*, 94 Tech. Rev. 28, 33 (Oct. 1991).

83. David F. Horrobin, *The Philosophical Basis of Peer Review and the Suppression of Innovation*, 263 JAMA 1438, 1440 (1990). This paper was cited by the majority in *Daubert* (1993), 509 U.S. at 593.

84. Horrobin, *supra* n. 83, at 1440.

85. *Id.*

86. *Daubert* (1991), 951 F.2d at 1131 n. 3.

87. *See Daubert* (1993), 509 U.S. at 593 (establishing that whether an idea has been subject to peer review is a "pertinent consideration").

88. Pressured to save money and publish on schedule, editors of peer-reviewed journals can sometimes publish the "scientific equivalent of a supermarket tabloid." Amici Br.

> Publication (which is but one element of peer review) is not a *sine qua non* of admissibility; it does not necessarily correlate with reliability . . . and in some instances well-grounded but innovative theories will not have been published . . . . But submission to the scrutiny of the scientific community . . . increases the likelihood that substantive flaws in methodology will be detected.[89]

It would have been desirable to have made the distinction between the broad and the narrow senses of "peer review" more explicit; nevertheless, what Justice Blackmun had in mind seems reasonably clear; moreover, it seems true: poor scientific work may pass pre-publication peer review, and good work may not, but when scientific work is published and made available for the scrutiny of other scientists, the likelihood increases that, eventually, any serious methodological flaws will be spotted. And Justice Blackmun's advice about the weight courts should give this "*Daubert* factor"—in effect, that it's a relevant consideration, but not necessarily a decisive one—seems at first blush quite unexceptionable:

> The fact of publication (or lack thereof) in a peer[-]reviewed journal thus will be a relevant, though not dispositive, consideration in assessing the scientific validity of a particular technique or methodology on which an opinion is premised.[90]

"At first blush"; but at second blush you find yourself beset by worries, both theoretical and practical: the meaning of "reliable" threatens to unravel into indeterminacy; and the Court's advice about the bearing of peer review on the determination of reliability sounds less and less helpful. Ambiguities strike one almost immediately: are courts to ask whether the work on which proffered testimony is based was published after surviving peer review, or is it enough that it be published in a "peer[-]review journal"? Should *the witness's* work have been subject to peer review

---

of Chubin, *supra* n. 7, at 12. A compromise between "absolute certainty" in the validity of the scientific claims in the articles and "absolute speed and absolute economy" means that "mistakes become inevitable and that erroneous, misleading, and fraudulent reports are sometimes published." *Id.*

89. *Daubert* (1993), 509 U.S. at 593.

90. *Id.* at 594.

and publication, or is it enough that the witness rely on *others'* peer-reviewed and published work? And so on.

Justice Blackmun's sense that survival of the long-run scrutiny of the scientific community is about the best indicator of scientific validity a layperson can have, albeit a fallible one, is perfectly correct; but it is of no real practical help. For obvious reasons the scientific issues at stake in legal cases are not likely to turn on the most firmly established science, but on the still-controversial stuff; and it would be hopelessly unrealistic to imagine that courts could somehow figure out which still-controversial scientific claims *will*, eventually, survive such "peer review," when scientists themselves cannot.

And rather than clarifying the concept of "evidentiary reliability" (which the *Daubert* Court equates with "scientific validity"),[91] Justice Blackmun's observations contribute to its obscurity.[92] In ordinary speech, "reliable" has a whole tangle of uses: but whether we are describing inanimate objects, like clocks or cars, or persons (also called "informants," or "sources"), or information, data-bases, etc., reliability—fitness to be relied upon—is ordinarily conceived as a matter of degree. But the *Daubert* ruling is about admissibility, which is *not* a matter of degree; and so obliges us to adopt a categorical conception.

If evidence must be reliable *enough* to be admissible, *how* reliable does it have to be, and how is a court to determine whether evidence meets the standard? (Is the same degree of reliability to be imposed on "soft" scientific evidence as on "hard," or on non-scientific expert testimony as on the scientific?) It seems to make sense, as Judge Becker argued in *In re Paoli Railroad Yard PCB Litigation*, that "[t]he evidentiary requirement of reliability [should be] lower than the merits standard of correctness";[93] for if the threshold for admissibility were as high as the standard of proof, a party seeking to introduce expert testimony would be re-

91. *Id.* at 591.

92. "Many factors will bear on the inquiry, and we do not presume to set out a definitive checklist or test. But some general observations are appropriate." *Id.* at 593. Under *Daubert* (1993), the key factors in determining evidentiary reliability, or scientific validity, are whether a theory or technique can be and has been tested, whether it has been subject to peer review, the known or potential rate of error, the standards controlling the technique's operation, and "general acceptance." *Id.* at 593–594.

93. *In re Paoli R.R. Yard PCB Litig.*, 35 F.3d 717, 744 (3d Cir. 1994).

quired, in effect, to prove his case twice—and the court would be trespassing on the jury's turf. But now you start to wonder: is peer-reviewed publication enough, after all, to guarantee that proffered evidence meets a minimal threshold standard of reliability? If not, is it at least enough to guarantee that, even if the conclusions drawn are unreliable, the methodology followed meets minimal standards? Isn't that what Judge Kozinski had in mind when he wrote in 1995 that peer-reviewed publication "is a significant indication . . . that it meets *at least the minimal criteria* of good science"?[94]

Justice Blackmun's ruling for the Court leaves all this open.[95] Chief Justice Rehnquist's ruling for the Court in *General Electric Co. v. Joiner*, casting doubt on the robustness of the distinction between methodology and conclusions on which *Daubert* had relied, is no help.[96] And Justice Breyer's opinion for *Kumho Tire Co. v. Carmichael*—holding that *Daubert* gatekeeping extends to nonscientific as well as scientific testimony, but that courts may use any, all, or none of the *Daubert* factors, and/or other factors more appropriate to the task at hand—confirms that the tricky stuff is to be left to courts' discretion.[97]

It is no surprise that the *Daubert* Court did not come up with a precise formula for deciding questions of evidentiary reliability; even if such a thing were feasible, it would probably be, not desirable precision, but the kind of "[d]elusive exactness" Oliver Wendell Holmes once decried as "a source of fallacy throughout the law."[98] And, especially given that "peer review and publication" is only one factor on *Daubert*'s flexible list, perhaps it is no surprise, either, to find no clear correlation of decisions to admit,

---

94. *Daubert* (1995), 43 F.3d at 1318 (emphasis added).

95. *Supra* nn. 91–94 and accompanying text.

96. *Gen. Elec. Co. v. Joiner*, 522 U.S. 136, 146 (1997) (asserting that "conclusions and methodology are not entirely distinct from one another").

97. *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 141 (1999). The Court held that

> *Daubert*'s general holding—setting forth the trial judge's "gatekeeping" obligation—applies not only to testimony based on "scientific" knowledge, but also to testimony based on "technical" and "other specialized" knowledge. . . . [A] trial court *may* consider one or more of the more specific factors that *Daubert* mentioned when doing so will help . . . . But . . . *Daubert*'s list of specific factors neither necessarily nor exclusively applies to all experts or in every case.

Id.

98. *Truax v. Corrigan*, 257 U.S. 312, 342 (1921) (Holmes, J., dissenting).

or to exclude, proffered expert testimony, and whether or not it satisfies this factor. Instead:

(1) some courts (citing Justice Blackmun's concession that peer-reviewed publication is not a sine qua non of admissibility) have admitted expert testimony not based on work that has been peer-reviewed and published;[99]

(2) some courts (citing Justice Blackmun's concession that peer-reviewed publication does not necessarily correlate with reliability) have excluded expert testimony based on work which has been peer-reviewed and published;[100]

(3) some courts (citing Justice Blackmun's acknowledgment that peer-reviewed publication is a pertinent consideration) have admitted testimony in part be-

99. *See e.g. Ruiz-Troche v. Pepsi Cola of Puerto Rico Bottling Co.*, 161 F.3d 77, 84 (1st Cir. 1998) (reversing the district court's exclusion of Dr. O'Donnell's testimony regarding the effects of cocaine on a driver's behavior, on the grounds that, although the secondary sources he cited were not peer-reviewed or published, other peer-reviewed, published studies made the same point); *Kannankeril v. Terminix Intl., Inc.*, 128 F.3d 802, 809 (3d Cir. 1997) (vacating and remanding the lower court's decision, which had excluded Dr. Gerson's testimony, arguing that "although Dr. Gerson did not write on the topic, his opinion is supported by widely accepted scientific knowledge of the harmful nature of organophosphates," and noting that *McCullock v. H. B. Fuller Co.*, 61 F.3d 1038, 1042 (2d Cir. 1995) held that peer review, publication, and general acceptance go to the weight, not the admissibility, of evidence); *Metabolife Intl., Inc. v. Wornick*, 264 F.3d 832, 843 (9th Cir. 2001) (citing *Daubert* (1995), 43 F.3d at 1317, for the proposition that "when research is begun pre-litigation, it may be reliable without peer review"); *U.S. v. Hankey*, 203 F.3d 1160, 1168 (9th Cir. 2000) (holding that the district court did not abuse its discretion in admitting the testimony of a police expert on gang codes and citing *Kumho*, 526 U.S. at 1176, saying a court must have latitude not only in deciding whether to admit expert testimony, but also in deciding "*how* to test an expert's reliability").

100. *See e.g. Allison v. McGhan Med. Corp.*, 184 F.3d 1300, 1313, 1316, 1319 n. 24 (11th Cir. 1999) (upholding lower court's exclusion of experts' testimony on role of silicone breast implants in causing the plaintiff's injuries, in part on the grounds that the fact that a study was peer-reviewed and published "does not mean it constituted an adequate basis" for experts' opinion, that "scrutiny by one's peers does not insure admissibility," and that the fact that a witness had published many articles in peer-reviewed journals "does not substantiate the scientific validity of his premise"); *U.S. v. Cordoba*, 194 F.3d 1053, 1059 (9th Cir. 1999) (upholding the district court's exclusion of polygraph evidence, even though hundreds of articles have been published on polygraphs, including many in peer-reviewed journals).

cause it was based on peer-reviewed and published work;[101] and

(4) some courts (also citing Justice Blackmun's acknowledgment that peer-reviewed publication is a relevant factor) have excluded testimony in part because it was not so based.[102]

Nor, given Justice Blackmun's shifts from broader to narrower senses of "peer review," is it altogether surprising that some courts have interpreted "peer review" to cover kinds of exposure to other people in a field other than pre-publication peer review.[103] Nor is it any surprise that "peer review and publication" has found its way into courts in states that have not adopted

---

101. *See e.g. In re Silicone Gel Breasts Implants Products Liab. Litig.*, 318 F. Supp. 2d 879, 896 (C.D. Cal 2004) (finding that Dr. Neugebauer's analysis and criticism of the existing epidemiological evidence is admissible, in part because "[t]he statistical underpinnings of epidemiology . . . have been subjected to peer review and publication").

102. *See e.g. Berry v. City of Detroit*, 25 F.3d 1342, 1350–1351 (6th Cir. 1994) (excluding testimony regarding police training of the plaintiffs' witness Leonard Postill, in part on the grounds that "[t]here certainly is no testimony as to any peer review of Postill's theory"); *Natl. Bank of Com. v. Associated Milk Producers*, 191 F.3d 858, 864–865 (8th Cir. 1999) (affirming the lower court's exclusion of expert's testimony as to connection between aflatoxin M-1 (AFM) and the plaintiff's cancer in part on the grounds that "[t]here are no scientific studies or medical literature that show any correlation between AFM and laryngeal cancer"); *Nelson v. Tenn. Gas Pipeline Co.*, 1998 WL 1297690 at **8–9, 13 (W.D. Tenn. Aug. 31, 1998) (excluding the testimony of Nelson's experts Drs. Kilburn and Hirsch that the plaintiffs' injuries were caused by PCB exposure from the gas pipeline, in part on the grounds that their work had not been published or peer-reviewed).

103. *See e.g. U.S. v. Bonds*, 12 F.3d 540, 559–560 n. 16, 568 (6th Cir. 1993) (affirming the lower court's decision to admit FBI's expert testimony on DNA, even though "many of the articles introduced as . . . exhibits did not appear in a 'peer-reviewed journal' in the strict sense of that term," since "all of the articles gave the FBI's procedures exposure within the scientific community"); *U.S. v. Havvard*, 117 F. Supp. 2d 848, 854 (S.D. Ind. 2000) (admitting the FBI's fingerprint-identification testimony, arguing and concluding that it satisfies *Daubert*; in particular, a fingerprint examiner's methods are subject to peer review because "any other qualified examiner can compare the objective information upon which the opinion is based and may render a different opinion if warranted"), *aff'd*, 260 F.3d 597 (7th Cir. 2001). *Havvard*, I believe, stretches the meaning of "peer review" well beyond all reasonable limits.

Daubert,[104] and even into cases involving quite different issues from questions of admissibility of expert testimony.[105]

But it is disappointing to find that courts' analyses of "peer review and publication" seem to have been, mostly, quite shallow. For our investigation of the virtues and vices of the pre-publication peer-review system has suggested a whole raft of questions that might throw light on the significance of the fact that the expert testimony proffered in a given case is, or is not, based on work published in a peer-reviewed journal. How epistemologically respectable is the field in question,[106] and are there serious ongoing methodological disagreements? Is this a highly regarded journal in the field, or a second- or third-tier publication—or a last resort of the desperate-to-publish? Was work published in a "peer-review journal" in fact peer-reviewed, or was it published by editorial privilege, or invited? If it was peer-reviewed, were the reviewers suggested by the author(s)? If it was invited, was this because of the author's good reputation, or because of his or her personal relationship with the editor? Is the author (or an author) associated with the journal, e.g. by serving on the editorial board? Does the journal in which the work was published receive support, direct or indirect, from one of the parties to the case or to closely related litigation? Was the work rejected by other journals before being accepted by this one, and if so, by how many, and which, and on what grounds? If testimony is based on work which has not been published, is that because it is too recent, or because, though not recent, it was never submitted for publication, or because it was submitted, but was rejected?

---

104. *See e.g. Berry v. CSX Transp., Inc.*, 709 So. 2d 552, 569–570 (Fla. 1st Dist. App. 1998) (arguing that even under *Frye* "[w]hile the existence of numerous peer-reviewed, published . . . studies does not guarantee that the studies are without flaws, such publication . . . alleviates the necessity of thorough judicial scrutiny . . . at the admissibility stage").

105. *See e.g. Kitzmiller v. Dover Area Sch. Dist.*, 400 F. Supp. 2d 707, 743–745 (M.D. Pa. 2005) (finding that, while the plaintiffs' experts' testimony is based on peer-reviewed literature, defendants' experts' testimony is not based on material that has been subject to peer review, which is "exquisitely important" in the scientific process, helping to ensure "that research papers are scientifically accurate[ ], meet the standards of the scientific method, and are relevant to other scientists in the field").

106. *See Kumho*, 526 U.S. at 151 ("[n]or . . . does the presence of *Daubert*'s general acceptance factor help show that an expert's testimony is reliable where the discipline itself lacks reliability").

Have there been subsequent expressions of concern or retractions,[107] or have other papers criticized the work?

These are not easy questions to answer, and it is not remarkable that courts have not routinely asked them. But when some of them *were* explored by a court—as it happens, in another Bendectin case, less well-known than *Daubert*—the results were instructive, to say the least; and quite disturbing.

V. FULL CIRCLE? "PEER REVIEW AND PUBLICATION" IN THE BENDECTIN LITERATURE

Blum v. Merrell Dow Pharmaceuticals[108] was a long, drawn-out Bendectin case from the Pennsylvania courts which began several years before *Daubert*, in 1982, but was not finally concluded until 2000.[109] (No, Pennsylvania has not adopted *Daubert*, but remains a *Frye* state; don't forget, however, that *Daubert* (1989) was a rare instance in which *Frye* had been used in a civil case, and that the Supreme Court granted certiorari to settle whether *Frye* had been superseded by the Federal Rules of Evidence.)

In *Blum* as in *Daubert*, Merrell Dow's attorneys argued that the plaintiffs' expert testimony should be excluded, on the grounds that it was not generally accepted in the relevant scientific community.[110] The Blums' attorneys argued, however, that Merrell Dow's expert testimony should be excluded, on the grounds that the supposed "scientific consensus" on the matter was completely artificial; that it had been created, in fact, by the defendant manufacturer's support of favorable research and—the key point for our purposes—by Merrell Dow's support of ques-

107. To find retractions in medical journals, visit PubMed (available at http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed&itool=toolbar) and search for "retracted publication" in the MeSH (Medical Subject Heading) database, then click on Links and select PubMed.

108. 764 A.2d 1.

109. *Blum*, 764 A.2d at 1 (final appeal decided December 22, 2000); *Blum*, 1 Pa. D. & C. 4th 634, 635 (Pa. Com. Pleas Ct. 1988), *aff'd*, 626 A.2d 537 (Pa. 1993) (lawsuit filed on September 13, 1982).

110. *Blum v. Merrell Dow Pharms., Inc.*, 705 A.2d 1314, 1319 (Pa. Super. 1997), *aff'd*, 764 A.2d 1 (Pa. 2000).

tionably peer-reviewed journals that would publish results helpful to the company in defending itself against Bendectin litigation.[111]

Judge Bernstein's ruling at the second trial in 1996 includes a devastating summary of the testimony of Merrell Dow's experts.[112] Defense expert Dr. Bracken acknowledges not only that articles that are "less than good" can pass peer review, but also that his own published study of Bendectin and birth defects was itself less than good.[113] Defense expert Dr. Klebanoff testifies that Bendectin does not cause birth defects, but then admits that his own article showed a statistically significant association with congenital cataracts, underdeveloped lungs, and microcephaly.[114] Defense expert Dr. Shapiro (whose unit at Boston University had received over one-and-a-half million dollars from Merrell Dow) testifies that a drug taken by a mother after the time of fetal limb formation could not cause a limb defect; but acknowledges under cross-examination that the data on which his opinion was based lumped together women who took Bendectin in the period in which fetal limbs were forming, and women who took it later, and that for this reason his study had been criticized in subsequent articles.[115] Defense expert Dr. Newberne, Merrell Dow's Vice-President responsible for animal testing and drug safety, testifies that while a study by Dr. Smithells concluding that Bendectin is not teratogenic was being reviewed and rejected by the *British Medical Journal*, *The Lancet*, and the NEJM, and finally accepted by a much less prestigious journal, *Teratology*, Dr. Smithells was actively soliciting funds from the company: "Much clearly depends upon the value of this publication to Merrell Dow," Dr. Smithells wrote, since it "may save the company large sums of money . . . in the California court . . . ."[116]

111. *See Blum*, 764 A.2d at 8 (Castille, J., dissenting).

112. *Blum v. Merrell Dow Pharms., Inc.*, 33 Phila. Co. Rptr. 193, 206–207 (Pa. Com. Pleas Ct. 1996).

113. *Id.* at 208. Dr. Bracken's study was based on interviews with 1,427 mothers, 122 of whom had taken Bendectin during pregnancy, and showed a statistically significant increased risk of birth defects when a mother used Bendectin and smoked. *Id.* at 207.

114. *Id.* at 208–209. On cross-examination Dr. Klebanoff testified that the positive association between Bendectin and clubbed feet, though not statistically significant, met the standard he used in his article for cataracts and vomiting; he agreed that an article that is "less than good" may pass peer review. *Id.*

115. *Id.* at 214–216.

116. *Id.* at 217, 219. Dr. Newberne also acknowledged that after a first study by Dr. Hendrickx found a statistically significant increase in heart defects in Bendectin-treated

And the editor of *Teratology*, Dr. Robert Brent,[117] who has been retained as an expert by Merrell Dow for eighteen years, testifies that his only formal education in epidemiology was one course in statistics, but that he considers himself a world authority on "secular trend data"—a field in which, Judge Bernstein comments, he is apparently the only practitioner.[118] Using his editorial prerogative to sidestep peer review,[119] he had published in his own journal an article entitled *Litigation-Produced Pain, Disease, and Suffering: An Experience with Congenital Malformation Lawsuits*, based on his review of depositions and trial transcripts, and concluding that seventeen out of seventeen plaintiffs lied;[120] and he had submitted a draft article entitled *Bendectin: The Most Comprehensively Studied Human Non-Teratogen, and the Foremost Tortogen-litigen* to Merrell Dow's attorney for editing, hoping to publish it in NEJM, JAMA, or *The Lancet*.[121]

Eventually, Merrell Dow prevailed. In 2000, the Pennsylvania Supreme Court ruled that while the Blums' expert testimony was arguably admissible under *Daubert*, "a standard somewhat less exacting than that of *Frye*," it was inadmissible under the *Frye* rule.[122] In dissent, however, Justice Castille reminded his colleagues that the trial court, citing *Frye*, had been impressed

monkeys, Merrell Dow funded a second study that arrived at results more favorable to the company. *Id.* at 221. The entry in Merrell Dow's financial records was: "Hendrickx' monkey study—defense." *Id.*

117. Dr. Brent's testimony was found to be incredible because "[his] testimony and manner suggested a degree of conviction in his own conclusions unwarranted in a discipline in which . . . explanations are only more or less probable." *Blum*, 764 A.2d at 10–11 (Castille, J., dissenting) (quoting *Wells v. Ortho Pharm. Corp.*, 615 F. Supp. 262, 291 (N.D. Ga. 1985)).

118. *Blum*, 33 Phila. Co. Rptr. at 223–224.

119. *See id.* at 224 (explaining that some of Dr. Brent's writings had been published in *Teratology* solely "because of his personal editorial prerogative").

120. *Blum*, 33 Phila. Co. Rptr. at 225; Robert L. Brent, *Litigation-Produced Pain, Disease, and Suffering: An Experience with Congenital Malformation Lawsuits*, 16 Teratology 1, 5 tbl. 1 (Aug. 1997).

121. *Blum*, 33 Phila. Co. Rptr. at 225. The article in question had in fact been published, though Judge Bernstein's ruling does not tell us this. *See* Robert L. Brent, *Bendectin: Review of the Medical Literature of a Comprehensively Studied Human Nonteratogen and the Most Prevalent Tortogen-Litigen*, 9 Reproductive Toxicology 337 (1995) (published version of Dr. Brent's article). The article prompted a lawsuit for defamation against Dr. Brent by Dr. Stuart Newman, who was one of the plaintiffs' experts in *Blum* and whom Dr. Brent misquoted. *Newman v. Brent*, 1998 U.S. Dist. LEXIS 10476 (D.D.C. 1998).

122. *Blum*, 764 A.2d at 3–4. Despite the *Blum* Court's rhetoric, however, it is questionable, to say the least, whether *Daubert* really *is* "less exacting" than *Frye*.

by the fact that "the scientific consensus [on Bendectin] derives largely from the proprietary influence and litigation interests of the adverse party," and that "much of the 'science' in this area, held up by Merrell Dow as the objective, generally accepted scientific view that requires exclusion of the plaintiffs' experts' 'contrary' conclusions, *itself* was a product of Merrell Dow's litigation-driven influence."[123]

But for our purposes, it is Judge Bernstein's conclusions that are most apropos: The testimony in this case, he observes, "demonstrates how 'scientific consensus' can be created through purchased research and the manipulation of a 'scientific' literature, funded as part of litigation defense, and choreographed by counsel."[124] It "clearly demonstrated that not all 'peer review' journals are created equal," that "not all the articles contained in 'peer review' journals were even reviewed . . . ," and that "[a]rticles were intentionally inserted in peer review journals for use in court."[125] And in Appendix B to his opinion, entitled "Science and Justice," Judge Bernstein adds the following:

> The testimony demonstrated medical-scientific peer[-]review journal literature created and manipulated for use in the courts . . . . The testimony demonstrated that articles were inserted in "peer review" journals, without review by independent authorities, but edited by lawyers . . . [and] revealed factual editing of supposedly scientific research literature by the very lawyers defending in litigation.[126]

* * *

The example is instructive, reinforcing Justice Blackmun's acknowledgement that peer-reviewed publication is no guarantee of "scientific validity" but at best a very fallible indicator; and reminding us that if courts were to pursue the questions suggested here, this *Daubert* factor could, and should, be handled with more caution, and more subtlety, than it has usually been up to now. It is also quite disturbing; for it suggests that the scientific litera-

---

123. *Id.* at 9, 11 (Castille, J., dissenting) (emphasis in original).
124. *Blum*, 33 Phila. Co. Rptr. at 230.
125. *Id.* at 246–247.
126. *Id.* at 248–249.

ture in the litigation by way of which "peer review and publication" entered the official legal vocabulary of admissibility may have been tainted by litigation interests. Ironically, it seems that the same commercialization of medical research that has contributed to the creeping corruption of peer review and publication may also have been partly responsible for the legal system's coming to rely on that process as a factor in determining evidentiary reliability.[127]

127. My thanks to Mark Migotti for helpful comments on two drafts; to Lee Tilson and Susan Shott for help in finding material on statistical errors in, and retractions of, published articles; and to the several librarians at the University of Miami Law Library, especially David Hollander and Barbara Brandon, who helped me find relevant material.