

JUDGING EXPERT TESTIMONY: FROM VERBAL FORMALISM TO PRACTICAL ADVICE

Susan Haack*

ABSTRACT: Appraising the worth of others' testimony is always complex; appraising the worth of expert testimony is even harder; appraising the worth of expert testimony in a legal context is harder yet. Legal efforts to assess the reliability of expert testimony —I'll focus on evolving U.S. law governing the admissibility of such testimony— seem far from adequate, offering little effective practical guidance. My purpose in this paper is to think through what might be done to offer courts more real, operational help. The first step is to explain why the legal formulae that have evolved over the years may seem reassuring, but aren't really of much practical use. The next is to suggest that we might do better not by amending evidentiary rules but by helping judges and attorneys understand what questions they should ask about expert evidence. I focus here on (i) epidemiological testimony, and (ii) the process of peer review.

KEYWORDS: evidence law; expert testimony; reliability; epidemiology; peer review; the *Daubert-Joiner-Kumho* régime.

In order that we may have the right to accept [someone's] testimony as ground for believing what he says, we must have reasonable grounds for trusting his veracity, that he is really trying to speak the truth so far as he knows it; his knowledge, that he has had opportunities of knowing the truth about this matter; and his judgement, that he has made proper use of those opportunities in coming to the conclusion which he affirms.

W. K. Clifford (1877).¹

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

¹ Clifford, 1947: 79 (first published in 1877).

If had to choose between these thoughts of Clifford's and the enormous mountain of literature on the "epistemology of testimony" produced over the last couple of decades, I'd take Clifford's briskly commonsense words any day. Just as he says, whenever we have to assess the trustworthiness of what another tells us, we must have some sense of our informant's honesty, his competence, and how well he has judged where *his* evidence points. But there are many layers of complexity here; so this is just a starting point.²

1. WHY IS ASSESSING EXPERT TESTIMONY SO PECULIARLY HARD?

Often, dealing with what others tell us seems quite simple. When we're consulting an airline representative about plane schedules, asking a passer-by for directions to the train station, or asking someone on the other end of the phone what the weather's like where he is, we hardly think about what is involved in figuring out whether, or to what degree, what we're told is credible. In fact, however, even in these familiar kinds of case it's quite a complicated business; it's just that we only notice the complexities when things don't go smoothly—when we're not sure the airline representative heard, or understood, our question correctly, when we're not sure an informant isn't deceiving us for reasons of his own, when we're not sure we followed our informant's rapid Spanish directions, when we're not sure whether our informant really knows himself, or is just guessing, ..., and so on. If we have to repeat our question several times, if the answer is given in a hard-to-understand accent or in a language in which we're not fluent, if our informant is hesitant or changes his initial answer ("go left at the church; no, sorry, right—no, sorry, I was right the first time, left"), if he may be trying to deceive us ("the robbers went that way!"), if it seems his vision may be less than perfect, ..., etc., we grow suspicious, and start wondering if his testimony can be relied on.

Still, in these everyday cases what our informant is telling us is nothing arcane or (linguistic snafus aside) hard for us to understand; and we have some idea what to look for in judging whether such an informant might have reason to mislead or misinform us, whether he is really in a position to know what he is telling us, and whether he might have jumped to a conclusion his evidence doesn't warrant. But when we are dealing with expert information, and our informant is telling us something technical that we are in no position to know for ourselves, judging his credibility is significantly harder.

To be sure, the line between everyday testimonial evidence and expert testimonial evidence is somewhat fuzzy. Perhaps, for example, I ask someone more expert than myself at word-processing how to ensure that my copyright notice appears as a footer on the first page of my paper but not on subsequent pages; I may not understand all

² See Haack, 2014a.

the computer jargon in his response, but I can probably easily enough check whether following his instructions gives the desired result. But if I ask an expert what is known about the age of the earth, what the paleontological evidence is for the theory of evolution, whether exposure to this chemical is a possible cause of someone's developing Lou Gehrig's disease, or even whether this design of seat-belt buckle will lock solid in a serious car-crash..., etc., though I'll no doubt understand the take-home message, I'll be very hard-pressed to understand the reasons for the expert's conclusion, or even to follow the technical language he will use.

Because it takes years of specialized work and study to become an expert on a certain subject, a lay person won't find it easy to assess an expert's competence, his judgment, or even, always, his honesty—after all, it's hard to spot implausibly exaggerated claims, or even inconsistencies, in testimony you only half-understand. Often you can do little more, besides looking for the familiar signs of dishonesty or uncertainty, than look to supposed experts' qualifications, their degree of confidence, the extent to which others in the field agree with them, and so on.

When we are assessing the worth of what witnesses say in a legal context, we need to take further complexities into account: that (normally) they speak on behalf of one side or another; that they will likely have been instructed not to volunteer information, but only to answer the questions they are asked; that they are bound by evidentiary rules that will make some relevant testimony inadmissible; and that—given that a great deal of money, or a person's freedom or even his life may be at stake—there may be strong emotions at work, encouraging witnesses to tell less than the whole truth, or even to lie outright.

And when it comes to evaluating the worth of expert testimony in a legal context we must deal, not only with the complications of appraising regular testimony in everyday contexts, the complications of appraising expert testimony in regular contexts, and the complications of appraising regular testimony in legal contexts, but all of these together; plus *further* complications. Experts hired by the parties will usually have been carefully screened and prepared to be as helpful as possible to their side; they may well have been drawn from the margins of those in their field—those most willing to reach a conclusion when others are unsure; their professional pride is likely to be at stake; and the scientific or other technical questions on which they testify may very well, as yet, be far from settled either way.

It's no wonder, then, that courts have such a hard time when they're faced with the unenviable task of determining whether expert testimony is reliable enough to be admitted, or whether it is sufficiently strong to establish a claim to the required degree of proof. They will need to master material presented in a technical jargon that may be far beyond their ken, and to make judgments about its relevance, its reliability and, in a bench trial,³ its weight—when experts in the field involved disagree among themselves. A federal judge once told me how she was struggling to

³ A trial where there is no jury, only a judge or panel of judges.

make sense of the expert testimony in a case involving a contested biotech patent: “What wouldn’t I give for a nice, simple curling-iron case!” Indeed; I’m sure I’d have felt the same way.

More to the present purpose, perhaps, it’s no wonder, either, that legal efforts to assess the reliability of expert testimony—I’ll focus in what follows on evolving U.S. law governing the admissibility of such testimony—seem so far from adequate, offering reassuring-sounding forms of words, but little effective practical guidance. Hence my sub-title, and my purpose, in what follows: to think through what might be done to offer courts more real, operational help. The first step, though, is to explain why I say that the legal formulae that have evolved over the years, though they may seem reassuring, aren’t really of much practical help.

2. HOW HAVE U.S. COURTS HANDLED THE PROBLEM?

U.S. law distinguishes the question of the admissibility of testimony (whether it may be heard by, or if it is heard but then challenged, whether it may be taken into account by, the finder of fact), from the question of its sufficiency to establish a given claim. Admissibility of evidence is regarded as a question of law; sufficiency of evidence is regarded as a question of fact. As we shall see, however, as U.S. law on the admissibility of expert testimony has evolved, there has been some blurring of this line.

For a long time—though there were already many complaints about expert witnesses, especially their willingness to say whatever the side that hired them wanted⁴—all that was required for admissibility was that the evidence be relevant and the proffered expert suitably qualified. Then in 1923 a federal appeals court in D.C.—asked to determine whether the testimony of an expert on a brand-new technique for determining whether a defendant was lying by tracking changes in his systolic blood pressure under questioning had been correctly excluded—ruled that such testimony was inadmissible: the technique was just too new and untried to be heard. The court went on to articulate the first legal test for the admissibility of novel scientific testimony—that the theory or technique on which it is based must be “suffi-

⁴ “[E]xperience has shown that opposite opinions of persons professing to be experts can be obtained in any amount.” *N.Y. & Erie R.R. Co.*, 62 U.S. 88, 101 (1858). “[The expert witnesses’ views] cannot fail generally to be warped by a desire to promote the cause in which they are enlisted. ... [Medical expert witnesses] come with such a bias on their minds to support the cause in which they are enrolled, that hardly any weight should be given to their evidence.” *Ferguson v. Hubbell*, 97 N.Y. 507, 514 (1884). “[Scientific witnesses] are selected on account of their ability to express a favorable opinion, which, there is great reason to believe, is in many instances the result alone of employment [in the case] and the bias growing out of it.” Himes, 1893: 409. “The position of an expert on the witness stand, who does not testify to what he has observed but merely expresses his opinion as to a situation or on facts which have been established by other witnesses, is anomalous in Anglo-Saxon law... [T]here is constant complaining and mistrust on the part of judges, juries, and lawyers of the expert witness.” Friedman, 2010: 247.

ciently established to have gained general acceptance in the particular field in which it belongs”;⁵ though it said nothing about how to identify, let alone appraise the soundness of, a scientific field. This test is known, after the name of the defendant, as the *Frye* Rule. It took a long while to catch on; but by the early 1980s it had been adopted by many states, and several large states still follow it to this day.⁶

In 1975, however, the common-law evidentiary rules were codified in the Federal Rules of Evidence. Rule 702, which applies to testimony regarding “scientific, technical, or other specialized knowledge,” said only that a qualified expert may testify if his testimony is relevant, and not otherwise excluded by law; it didn’t mention *Frye* or “general acceptance.” For many years, it was unclear whether FRE 702 had, or hadn’t, superseded *Frye*; until in 1993 the U.S. Supreme Court ruled in *Daubert*—an otherwise routine case against Merrell Dow for its morning-sickness drug Bendectin, in which, most unusually, *Frye* had been used in a civil case—that it had.⁷ Federally at least, *Frye* was gone.

Nevertheless, Justice Blackmun’s ruling continued,⁸ though the new régime was more hospitable to the admissibility of expert testimony than the old, courts still had an obligation to screen expert testimony not only for relevance, *but also for reliability*—which, as I hinted earlier, shifted the line between admissibility and weight somewhat. To determine whether expert testimony is reliable enough to be admitted, Justice Blackmun argued, they should not look to the expert’s conclusions, but to his methodology—presumably so that the admissibility inquiry not trespass too far onto the turf of the fact-finder. What courts should ask, according to Justice Blackmun, is whether experts’ methodology is genuinely “scientific”; and to help them do this, he offers a “flexible list” of factors they may consider: can [the technique or method on which the testimony is based] be tested, and has it been? Has it been subject to peer review and publication? What is the known or potential error rate? And (in a nod to *Frye*) has it gained widespread acceptance in its field?

In 1995, when *Daubert* was finally decided on remand to the Ninth Circuit, Judge Kozinski added a new factor to Justice Blackmun’s list: Was the work on which the testimony is based litigation-driven, or was it conducted independently of litigation?⁹

⁵ *Frye v. United States*, 293 F. 1013, 1014 (D.C. Cir. 1923).

⁶ The story about Florida law on this matter is too extraordinary not to include. Florida first adopted *Frye* in *Kaminski v. State*, 63 So.2d 339, 340 (1952). For decades afterwards, it followed. *Frye*—mostly, except when it didn’t. In 2013 the Florida legislature voted to change to *Daubert*, and the governor signed off on the change. See 2013 Fla. Laws 107 (codified as Fla. Stat. § 90.702 (2018)). But the change is procedural, and so the last word rests with the Supreme Court of Florida. In late 2018, a decision of this court rejected *Daubert* and stuck with *Frye*. *DesLisle v. Crane Co.*, 258 So.3d 1219 (Fla. 2018). (Fla 2019). But in May 2019, the same court (though with three new members) ruled that Florida now is, after all, a *Daubert* state! In re: Amendments to the Florida Evidence Code, No.SC19-107 (Fla. May 23, 2019).

⁷ *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 587 (1993).

⁸ *Ibid* 594-95.

⁹ *Daubert v. Merrell Dow Pharm., Inc.*, 43 F.3d 1311, 1317 (9th Cir. 1995). Notably, Judge Kozinski makes an exception in the case of forensic testimony. *ibid* 1317 n.5.

Only a few years after constructing the elaborate apparatus of *Daubert*, however, the Supreme Court began *deconstructing* it. *G.E. v. Joiner* (1997) primarily concerned the standard of review of exclusions of expert testimony; which, the Court ruled, remained abuse of discretion. But—probably because *Joiner*’s attorneys had argued on appeal that, since their experts had used the *same* methodology as G.E.’s, it had been a legal error to exclude their experts and admit their opponents’—the *Joiner* ruling also quietly repudiated the distinction between methodology and conclusions that had been central in *Daubert*. Courts could look to experts’ conclusions as well as their methodology, Justice Rehnquist wrote for the majority; and expert testimony could be ruled inadmissible should the judge deem the “analytical gap” between data and conclusions too great.¹⁰

And only a couple of years after *Joiner*, faced with a case that obliged it to determine whether *Daubert* (where all the expert testimony involved had been scientific) also applied to *non*-scientific experts, the Supreme Court qualified the role of that “flexible list” of factors that courts might consider in determining whether expert testimony was reliable enough to be admitted—to the point where everything was, in effect, left to courts’ common sense. Where non-scientific experts like the expert on tire design in this case were concerned, Justice Breyer wrote for the majority in *Kumho Tire* (1999), these factors might or might not be relevant; indeed, he continued, they wouldn’t all always be helpful even where the expertise concerned *was* scientific. So courts should use any, all, or none of those factors, or such other factors as might seem appropriate to the case. And he warned—improving, in this respect, on *Frye*—that they should bear in mind, when considering whether some theory or technique is widely accepted in the relevant field, that not all fields are equally sound, and some may be downright illegitimate; in which case, of course, widespread acceptance in the field is no assurance whatever of reliability.

The following year, FRE was revised; the new rule, requiring that the testimony be based on *sufficient* data, *reliably* arrived at by *reliable* methods, didn’t reflect the *Daubert-Joiner-Kumho* trilogy very closely; but it *did* reflect the fact that these decisions, rather than making it easier to get expert testimony admitted, had in fact made it harder. This 2000 version of the Rule remains in force.

I don’t suppose anyone imagines that the *Daubert-Joiner-Kumho* trilogy and the revised FRE 702, or the courts’ efforts to apply them over the years—though they have doubtless taught us some useful lessons—have entirely resolved the core problems with expert testimony, or even very significantly ameliorated them. Part of the trouble, of course, is that, while it’s clear that the underlying idea is to ensure that fact-finders aren’t misled by unreliable expert testimony, it’s completely *unclear* how reliable is reliable *enough*. But this is not a problem that can be solved by legal *epistemology*, however sophisticated. It calls, rather, for a legal *decision*, presumably

¹⁰ Gen. Elec. Co. v. Joiner 522 U.S. 136, 146 (1997). See also Haack, 2016.

informed by policy considerations: e.g., that the standard is higher than “scintilla,” or than “reasonable suspicion,” but lower than “clear and convincing.” Another part of the trouble, however, is that these standards of admissibility are ambiguous between formulae which courts can apply relatively straightforwardly, but which correlate only very weakly with reliability, and formulae that better correlate with reliability, but that courts are understandably at a loss to put into operation. We see this with *Frye*, and all through the various standards since.

The first phrase of the *Frye* standard, “sufficiently established to be generally accepted in the field to which it belongs,” has almost always been dropped,¹¹ and the Rule truncated to “generally accepted in the field to which it belongs.” No wonder: this is a lot easier for a lay judge than deciding how well-established a scientific claim is. But the *Frye* Rule tells courts nothing about how to determine what the relevant field is; nor, as mentioned, does it address the problems raised by the possibility that an entire field (e.g., forensic bite-mark identification)¹² may be flaky; and it is obviously susceptible to the guild problem, the danger that a group of practitioners in an iffy field will readily testify as to the general acceptance of their techniques. So the impression that *Frye* really gives judges a simple, operational way to decide whether novel scientific evidence is reliable enough to admit is a kind of illusion.

The original, 1975, version of FRE 702 says nothing explicit about reliability; and appears on its face to be very hospitable to the admission of expert testimony. Arguably, however, it might be thought implicit that such testimony should, in accordance with FRE 403b, be reliable enough that hearing it will help, rather than confuse or mislead, the finder of fact. At any rate, the Supreme Court’s ruling in *Daubert*, interpreting FRE 702 as formulated in 1975, finds the requirement that expert testimony be reliable implicit in the requirement that it be helpful to the finder of fact. But it’s quite clear that, whatever the Supreme Court’s intention, *Daubert* has proven less, not more, hospitable to the admissibility of expert testimony than *Frye* and than the original FRE 702, which it supposedly interprets.

The *Daubert* ruling is more complex, and in a sense much more sophisticated, than *Frye* or FRE 702; and for that very reason, it has proven far more expensive in terms of experts’, attorneys’, and judges’ time. Unfortunately, however, it confusedly identifies *reliable* expert testimony with *scientific* expert testimony—when obviously not all, and not only, scientific expert testimony is reliable. Most to the present purpose—and even if the distinction between methodology and conclusions were easily applicable, which it surely isn’t—those “*Daubert* factors” are much less operationally effective than they might at first appear.

The first *Daubert* factor (“can [the testimony] be and has it been tested?”), rests on a confused *mélange* of Karl Popper’s and Carl Hempel’s incompatible philosophies of science.¹³ It might be helpful in, for example, a case of knife-mark identi-

¹¹ Not always, however. See *Ramirez v. State*, 810 So. 2d 836, 843 (Fla. 2001).

¹² See Beecher-Monas, 2009.

¹³ See Haack, 2005.

fication testimony that could easily enough be tested, but hasn't been. More often, though, it has understandably proven simply baffling to courts.¹⁴ The second, which Justice Blackmun himself acknowledges is “not dispositive,” is ambiguous: are courts to look to whether the relevant work has passed pre-publication peer review—relatively easy to determine, but only a very weak indication of reliability—or to whether it has survived, or will survive, the long-run scrutiny of the scientific community after publication—a much better indicator of reliability, but impossible for judges to determine.¹⁵

The third *Daubert* factor (“known or potential error rate”), though certainly pointing to something one would want to know when deciding whether a technique or test is good enough to admit, is too lacking in specificity to be very helpful: what level of error rate is too high, one wonders, and what exactly is the force of that “potential”?¹⁶ And shouldn't courts also look to the *kind* of error to which a technique is susceptible: to false positives or false negatives, for instance? And the last factor, of course, is essentially a slightly weakened version of the *Frye* Rule, with all its troubles.

The new *Daubert* factor introduced by Judge Kozinski—is the science litigation-driven?—suffers the same problem as the first two of the Supreme Court's original factors: it's relatively easy for a court to determine whether scientific work was undertaken before, or after, litigation had begun, but this has little bearing on its reliability; while whether the work was motivated or otherwise biased by the need to find evidence helpful to litigants, while a better indicator of reliability, is often nearly impossible to determine.¹⁷

Joiner takes away the methodology/conclusions distinction, a central plank of the *Daubert* platform; and as a result opens the door to admitting one party's experts while excluding the other's even when they employ the *same* methodology.¹⁸ And it replaces the methodology/conclusions distinction by something even more legally-formulaic, the talk of “analytical gaps” between data and conclusions.

Kumho Tire, however, is a different kettle of fish. It takes the deconstruction of *Daubert* a step further, making that “flexible list” of factors to consider *so* flexible as to leave almost everything to courts' discretion: they may use any, all, or none of them, or other factors they deem more appropriate. At the same time, however, it includes much good sense; for it's perfectly true, as Justice Breyer writes, that not every

¹⁴ Haack, 2010. See also *United States v. Bonds*, 12 F.3d 540, 552 (6th Cir. 1993) (arguing that, since the FBI's DNA work has been tested and found to be unreliable, testimony of its results should be admitted—as reliable!).

¹⁵ See Haack, 2007.

¹⁶ I suspect that Justice Blackmun may be recalling *Barefoot v. Estelle*, where admitting psychiatric testimony that Mr. Barefoot would be dangerous in future was held to be acceptable even though an amicus brief from the American Psychiatric Association reported that such predictions are wrong two times out of three, 463 U.S. 880, 899 (1983), and he wrote a passionate dissent. *ibid* 916.

¹⁷ See Haack, 2008.

¹⁸ Of course, it's not easy to say when two methodologies are the same, and when different.

factor (such as “peer-review and publication”) will be relevant to every instance even of scientific testimony, much less of non-scientific expert testimony; and perfectly true, too, that consensus in a poor field is no indication of reliability, so that courts must pay attention to the legitimacy of the field in question.

The revised FRE 702 (2000) is a little more helpful than the original insofar as it indicates that courts are to look at the sufficiency of the data, the methods by which the data were arrived at, *and* the application of the theory or technique; but its repeated calls for “reliability” and “sufficiency” tell courts nothing specific about how to determine this, leaving an impression of ritual incantation rather than practical advice.

In 1994, in the wake of *Daubert*, the first edition of the *Reference Manual on Scientific Testimony* was issued by the Federal Judicial Center; a second edition appeared in 2000, and a third in 2011.¹⁹ As with any team-written work, the various chapters of the book are variable in quality; some are weaker, but others are very competent statements of the state of the art in various scientific fields that are likely to come up in court with some frequency. I’m sure it has its uses; I’m less sure, however, that it has much to offer in the way of practical advice either for courts determining the reliability of expert scientific testimony, or for attorneys preparing to mount, or to defend against, a *Daubert* challenge. The 2009 NRC report, *Strengthening the Forensic Sciences in the United States*²⁰ (which of course is narrower in scope than the *Reference Manual*) offers more in the way of practical advice, but this is primarily advice about how forensic techniques and forensic labs might be improved, not advice about how courts might better appraise forensic evidence. In short, courts aren’t given as much practical help as they might wish in handling the formidable responsibilities *Daubert* requires of them if the admissibility of expert testimony is challenged.

3. HOW MIGHT WE DO BETTER?

But doing better isn’t easy: the task is a formidable one.

Courts have to determine whether expert testimony is reliable enough to be admissible in circumstances where experts in the field in question can’t agree about what’s reliable and what isn’t. This, however, is not so much a problem to be solved as a simple, unavoidable fact. Moreover, when they are applying *Daubert*, courts must decide whether the proponent has shown, *by a preponderance of the evidence*, that their expert testimony is *reliable enough to be admitted*—a double-barreled legal formula the meaning of which is, to say the least, far from transparent. But what this problem requires is close legal attention to that key word, “enough.” On top of this, though, the sheer variety of kinds of expert testimony with which courts may be faced means that the kind of very generic advice the Supreme Court has offered

¹⁹ Federal Judicial Center, 1994; Federal Judicial Center, 2000; Federal Judicial Center, 2011.

²⁰ National Research Council, 2009.

may not always be apropos. And this suggests a place to begin my practical-epistemological reflections.

Some kinds of issue arise over and over with respect to a wide variety of different kinds of expert testimony: issues about the soundness of statistical evidence proffered, for example, or the relevance of peer-reviewed publication to the likelihood that proffered scientific conclusions will prove solid. Other kinds of issue, though specific to certain kinds of testimony, arise many times because this kind of testimony is so frequently proffered and presented: issues about DNA identification evidence in criminal cases, for example, or epidemiological evidence in toxic torts. To deal with all of this would take at least a fat book or maybe, more realistically, a whole series of books; here, I can only make preliminary suggestions about a couple of issues—one in each of the two large categories just mentioned—to which I have given a lot of thought: (i) the kinds of epidemiological evidence frequently proffered in toxic-tort cases and (ii) the matter of peer-review and publication. My strategy will be to suggest the kinds of question those concerned to evaluate the reliability of expert testimony might reasonably ask, and to indicate what answers are to some degree favorable, what to some degree unfavorable, and what neutral with respect to the reliability of the testimony concerned.

3.2 Epidemiological Testimony

Like legal proof, the reliability of expert testimony isn't a matter of "yes" or "no," but always a matter of degree, of more and less; and what I can offer will be only indicia, not "criteria"—the *Daubert* Court was right to insist that its list of factors could not be used as a checklist, but must be treated as flexible.

Epidemiological testimony will be to the effect that a correlation, an increased relative risk, has, or hasn't, been found, between exposure to some substance (the alleged toxin at issue in the case) and some disease or disorder (the alleged disease or disorder the plaintiff claims to have suffered); it may, or may not, also include some more speculative statements about a possible causal connection. Here are some questions it might be helpful to ask in determining how reliable such testimony is likely to be.²¹

Is the disease or disorder alleged by the plaintiffs well-recognized and well-understood in the relevant medical field—or is it more like the "smoldering" Guillain-Barré Syndrome, a hitherto unheard-of form of the disease, that Mr. Manko claimed was caused by his swine-flu vaccination²² or, like the conglomeration of connective-tissue disorders claimed to have been caused by silicone breast-implants, a new congeries of disorders? If the existence of the disease/disorder/syndrome is contested in court, does this seem to be because one party's expert is adapting standard definitions to suit his side, or is it a matter of honest but as yet unresolved disagreement in the relevant field of medicine?

²¹ I draw in part, here, on an earlier paper, Haack, 2018.

²² *Manko v. United States*, 636 F. Supp. 1419, 1427-28 (W.D. Mo. 1986).

Sometimes, there is legitimate disagreement within a medical field about what mix of symptoms constitute a disease or disorder; and sometimes, the consensus on how best to characterize a disorder shift over time. One reason such a big scare was caused by (very weak) evidence that there might be a connection between MMR vaccine and autism, for example, was that diagnoses of autism shot up after this vaccination was made compulsory; not, however, because there were more cases of autism in the older sense, but because the medical community had adopted a new, broader definition of “autism-spectrum disorders.”²³ That the plaintiff’s disorder is well-defined and understood is a favorable sign; that there’s legitimate disagreement about its definition is neutral; that there seems to be disagreement only in litigation contexts where one side is trying to stretch or shrink a standard definition is unfavorable.

Are the alleged injuries supported by medical diagnoses, or are the diseases or disorders self-reported?

That the alleged injuries are self-reported, as in the Henneckens Study²⁴—the only study to find a correlation between silicone breast-implants and connective-tissue disorders—is an unfavorable sign; the more so if, as in this instance, there has been a good deal of publicity about the possible side-effects of a drug, a medical device, or a pollutant.

Is the substance involved in the study the same as that at issue in the case, a variant of it, or something chemically similar? Were the doses in the study similar to those suffered by the plaintiff(s) in the case? What about the manner of delivery (ingestion, inhalation, etc.)—was that the same or different from in the instant case?

That the same substance was involved is a favorable sign; that the substance was different is unfavorable, to a degree depending on how different it was; that the mode of delivery was the same is a favorable sign, that it was different, unfavorable. The fact that the studies cited by Joiner’s experts (at least, those still in the record by the time the case came to the Supreme Court) involved different substances, sometimes in significantly larger doses than Mr. Joiner’s exposure, and delivered differently, was doubtless why the Court thought they were at best indirectly relevant, and not reliable evidence in support of Joiner’s claim.

How large was the sample size? How probable is it that a correlation was found by chance, simply because the sample was too small?

I am deliberately avoiding the phrase “statistically significant,” both because it is so often confused with simple significance or importance, period, and because the level of statistical significance deemed acceptable is purely a matter of convention, set high by epidemiologists so as to avoid sounding false alarms. But of course, the larger the sample is, then, other things being equal, the better the study. Andrew Wakefield’s dreadful work supposedly finding a correlation between MMR vaccina-

²³ According to the standards of the 1992 edition of the *Diagnostic and Statistical Manual of Mental Disorders* (published by the American Psychiatric Association), 1 in 68 children had autism; while according to the standards of the 1994 edition, the figure was 1 in 59. McGinty, 2018.

²⁴ Henneckens, et al., 1996.

tion, bowel disorders, and autism—based on a sample of only 12 children²⁵—is a paradigm example of a bad study.

Was the study reported appropriately designed? Was it properly conducted? Was it randomized, controlled, double-blinded? Did it use the statistical techniques it said it used, and did it apply them correctly? Might those techniques (e.g., some programs for meta-analysis of multiple studies) potentially introduce bias?

The more serious the problems with the design and/or the conduct of a study, obviously, the less solid and reliable the results it reports. I have mentioned only some of the most obvious pitfalls; but other design faults, though harder to spot, are well within the reach of a layperson—such as the Seattle pharmacist who noticed that Merck’s VIGOR study of Vioxx underestimated the likelihood of adverse cardiovascular effects by tracking them for a shorter period than gastrointestinal effects, and many times urged the *New England Journal of Medicine*, which published the study, to issue a warning.²⁶

To be sure, even a not-very-well designed or not-very-well conducted study *may* uncover a real correlation. But we need to be alert to the possibility that the flaws in the design or the conduct of a study are such as predictably to skew the results: as the failure of some of Merrell Dow’s studies of Bendectin to distinguish women who took the drug during the period of pregnancy when limbs were forming and those who took it before, or after, that period predictably underestimated the relative risk of limb-reduction birth-defects in their children.²⁷

In this context we may start to ask about the purpose for which the study was undertaken:

Do those who paid for or conducted a study have an interest in reaching a given conclusion (were they, for example, scientists working for manufacturers hoping to establish that their medication is effective and safe, or were they scientists working, like Wakefield, with attorneys for one party or another)?

That those conducting a study had an interest in its coming out this way rather than that is, of course, an unfavorable sign—though not, to be sure, necessarily fatal.

Were there other such studies? Were those other studies published? If other studies were conducted, were they unfavorable, or less favorable, to the product than the published studies?

There *might* be epistemologically good reasons for not publishing a study (e.g., that as you were writing it up you realized there was something wrong with the design). But if there are other less-favorable studies that went unmentioned or unpublished, and especially if this is all work conducted or paid for by a manufacturer

²⁵ Wakefield, 1998.

²⁶ Armstrong, 2006.

²⁷ Blum v. Merrell Dow Pharm., Inc., 33 Phila. Cnty. Rep. 193, 214-17 (1996).

with an interest in getting FDA²⁸ approval to market his drug or in finding information useful in advertising, we have reason for caution about how reliable the published study is. The more serious the problems with the design and/or the conduct of a study, obviously, the less solid and reliable the results it reports.

But, before we turn our attention to issues about publication more generally, we should note—remembering that epidemiological studies can only show correlations, not causation—that there are further questions that judges and attorneys should ask if an expert witness offers epidemiological studies to establish causation.

*Is there further evidence to indicate that the correlation is not, e.g., the result of some common cause both of exposure to the substance and the development of the disorder, but is a matter of the exposure's causing the disorder? Does the witness suggest what causal mechanism might be at work? If so, what independent evidence is there of its biological soundness? Does he relate this postulated causal relationship to other, well-established biomedical knowledge? Does he refer to the factors that Austin Bradford Hill famously suggested decades ago? And if so, is he aware that these are only indicia, nothing like "criteria," or a checklist?*²⁹

If the witness simply *assumes* that a correlation is sufficient to show causation, that's a very bad sign; but the more, and the stronger, the reasons he offers for thinking that, in the present instance, there is a causal relationship, the better sign it is.

We might also look, as the courts involved did in *Daubert*, following Merrell Dow's lead, at whether the epidemiological studies cited have been published in peer-review journals, or only in court reporters' accounts of cases. This leads us directly to the second topic on which I'll try to offer some practical advice, issues about peer-review and publication.

3.2. Peer Review and Publication:

A necessary preliminary to assessing the weight that peer-reviewed publication should carry is to distinguish fields where such publication is usual and expected from others where it isn't. In an academic field, such as epidemiology or toxicology, looking to whether an expert witness's publications, or the publications on which an expert witness relies, are peer-reviewed makes sense; in fields where it's more a matter of a technique or specialized skill, such as police-training techniques³⁰ or the proper design of folding lawn-chairs, it is unrealistic to expect such publications.

²⁸ "FDA" abbreviates "Food and Drug Administration," referring to the body that regulates drugs, medical devices, etc., in the U.S.

²⁹ Hill, 1965; Haack, 2014b.

³⁰ *Berry v. City of Detroit*, 25 F.3d 1342, 1347 (6th Cir. 1994).

Supposing questions about peer-review *are* apropos, though, there's a good deal one would like to know before deciding that having survived this process is any indication of reliability; for example:³¹

If the work concerned is in a field where peer-reviewed publication might reasonably be expected, but hasn't been published in a "peer review journal,"³² why is this? What it completed too recently to have been submitted yet? Or was it submitted but rejected—and if so, by what journal(s)?

That work too recent yet to have been submitted for publication and refereeing hasn't been peer-reviewed isn't in itself any indication of unreliability; though of course very recent work hasn't had time to prove itself, and might turn out to be unreliable. That work has been submitted but rejected isn't a good sign, but isn't necessarily a strong indicator of unreliability, either; it depends what journal(s) rejected it, and why. Some journals have very high rejection rates, others much lower. If the *New England Journal of Medicine (NEJM)* rejects an article on a restricted-interest topic by a young, unknown scientist at an undistinguished university, that would hardly be surprising; but it wouldn't necessarily be a sign that the work is flawed.

Even if the work is published in a peer-review journal, was it in fact peer-reviewed, or was it published by invitation? If by invitation, was this because the author is a well-known leader in the field, or because he's a friend of the editor's, or a member of the editorial board?

That work appears in a "peer-review journal," by itself, means nothing; many such journals also publish invited papers, and/or papers given at symposia whose organizers have paid a substantial publication fee. But invited papers aren't necessarily less reliable than invited ones. They *may* represent important work by major figures in the field; on the other hand, they *may* be weak efforts by the editor's cronies, or simply papers that reflect the editor's prejudices.³³

If there are many authors, what role did each play in the work? Are any of those listed "ghost" writers whose names are there simply to add prestige,³⁴ or attorneys involved in relevant litigation?

Most scientific articles today have many authors; and in some instances—as with Dr. Wakefield's paper on MMR vaccine and autism—some names may appear, not because these people really had much, or anything, to do with the work, but to lend an appearance of solidity to the paper. This is never a good sign; and when, as with

³¹ I refer readers to *Blum v. Merrell Dow* (1986), where Judge Bernstein's ruling reveals that Merrell Dow's support had created a body of peer-reviewed literature indicating that it was generally accepted that Bendectin was harmless. See also Haack, 2007.

³² Because not all articles published in self-described "peer-review journals" are peer-reviewed: some are invited, others symposium papers the publication of which is subsidized organizers.

³³ Dr. Robert Brent, who served for many years as expert witness for the defendants in the Bendectin cases against Merrell Dow, also served for many years as editor of *Teratology*, where some of the papers finding Bendectin harmless were published. See *Blum v. Merrell Dow Pharm., Inc.*, 33 Phila. Cnty. Rptr. 193, 222-28 (1996).

³⁴ As, it seems, were those listed as co-authors of Wakefield's paper.

some of Dr. Brent's work on Bendectin, the "co-authors" are actually attorneys working for one party in the litigation, it is a very bad sign indeed.

*If an article was peer-reviewed, was the author (or were the authors) asked to suggest referees themselves?*³⁵ How many referees were consulted? Were any of them colleagues or friends of the authors? How long were they given to turn in their reports? Were they given access to the authors' raw data? Were they asked to check any statistical calculations in the work?

It's apparently routine at medical journals for authors to be asked to suggest referees; so, while one might well feel that this in itself threatens to introduce conflicts of interest,³⁶ this can't be treated as disqualifying. And mostly, as I understand it, referees are given only a short time to report back, don't have access to the raw data, and don't check statistical calculations. These, of course, are some of the reasons why the fact that an article was peer-reviewed is at best only a very weak indication of its quality.

What is the status of the journal concerned? Is it highly ranked, or a journal of last resort for those who must publish something to get tenure?

To be sure, there's an element of self-fulfilling prophecy about journal rankings (e.g., people want to publish in the highly-ranked journals, so these have higher rejection-rates, which in turn feeds into their high ranking). Still, by and large and on the whole, the quality of a publication in a "prestigious" journal is likely to be better than a publication in the *Journal of Desperation*. But even publication in the most prestigious forum is no guarantee: witness the badly flawed study of Vioxx published in the *NEJM*, the premier English-language medical journal.

How is the journal supported financially?

The fact that a defendant manufacturer supports a journal, for example, or advertises heavily in it, would be an unfavorable sign; as would be the fact that a journal clearly represents only one view in a field where opinion is split.

Has there been any subsequent "Expression of Concern" about, or retraction of, the work at issue?

Expressions of concern and, especially, retractions are obviously bad signs; and so, equally obviously—though this will take serious work to discover—is the reliance of work cited by an expert on other work that has been retracted.

Has the work been widely cited? If so, is this because it's so important it has attracted others' attention, or because it's so weak it's an easy target of criticism?

As the second part of the question indicates, numerous citations *may* be a good sign, but aren't always so.

³⁵ A common practice at medical journals. See Haug, 2015.

³⁶ And has, we know, led to outright fraud on the part of some Chinese medical scientists. *ibid.*

4. WHAT WOULD BE THE “PARTICULAR GO” OF THIS ADVICE?

My proposals are not intended as an alternative to the *Daubert* trilogy; nor are they meant to suggest how to conduct any and every *Daubert* hearing involving these issues. The *Daubert* régime arguably already absorbs too much of judges’ and attorneys’ time and energy,³⁷ and I don’t want to make this task more burdensome than it already is.³⁸

What *would* be practically helpful, I believe, is for judges and attorneys to have available a book, or series of books, regularly updated, suggesting what questions it might be advisable to ask and what answers are favorable, and what unfavorable, to the reliability of expert testimony, should this or that kind of issue arise. The hope is that some courts and some lawyers will ask those questions and draw reasonable conclusions from those answers; and that other attorneys and other courts will gradually, over time, learn from their experience, ..., and so on. The idea, if you like, is to provide the means gradually to nudge legal players in the direction of identifying better experts, and undertaking better-informed scrutiny of expert testimony. This won’t be a panacea; there *can be no* simple, one-shot solution to such a complex and multi-faceted problem. But it could be one useful step in the direction of improving both the quality of expert testimony, and courts’ appraisals of its worth.³⁹

BIBLIOGRAPHY

- Armstrong, D., 2006, May 15: «How the New England Journal missed warning signs on Vioxx: Medical weekly waited years to report flaws in article that praised drug», in *Wall Street Journal*: <https://www.wsj.com/articles/SB114765430315252591>
- Beecher-Monas, E., 2009: «Reality bites: The illusion of science in bite-mark evidence», in *Cardozo Law Review*, 30 (4): 1369-1410.
- Capon, R., 2016: «The performance of the Italian civil justice system: An empirical assessment», in *Italian Law Journal*, 2 (1): 15-31.
- Clifford, W.K., 1947: «The ethics of belief» (1877), in *The ethics of belief and other essays* (eds. Stephen, L., & Pollock, F). London: Watts & Co., 70-96.
- Federal Judicial Center, 1994: *Reference manual on scientific evidence*. Washington, D.C.
- 2000: *Reference manual on scientific evidence*. Washington, D.C.
- 2011: *Reference manual on scientific evidence*. Washington, D.C.
- Friedman, L. M., 2010: «Expert testimony, its abuse and reformation», in *Yale Law Journal*, 19 (4): 247-257.

³⁷ One reason, I suspect, why the Florida Supreme Court initially declined to endorse the legislature’s shift to *Daubert*. See Hal, 2016 (reporting the former head of the Florida Bar Association as saying that the shift to *Daubert* was taking up far too much of judges’ time); and *DeLisle v. Crane Co.*, 258 So.3d 1219, 1233 (Fla. 2018) (Pariente, J., concurring), making the same point.

³⁸ I think here of the troubles of the Italian Civil Justice System: notoriously slow, and made even slower by some attempts at reform. See e.g., Capon, 2016.

³⁹ My thanks to Mark Migotti for helpful comments on a draft, and to Nicholas Mignanelli and Bianca Anderson for help with footnotes, references, and formatting.

- Haack, S., 2005: «Trial and error: The Supreme Court's philosophy of science», in *American Journal of Public Health*, 95 (Supplement 1): S66-73. Reprinted as «Trial and error: Two confusions in *Daubert*» in Haack, 2014a: 104-21.
- 2007: «Peer review and publication: Lessons for lawyers», in *Stetson Law Review*, 36 (3): 789-819. Reprinted in Haack, 2014a: 156-79.
- 2008: «What's wrong with litigation-driven science? An essay in legal epistemology», in *Seton Hall Law Review*, 38 (3): 1053-83. Reprinted as «What's wrong with litigation-driven science? » in Haack, 2014a: 180-207.
- 2010: «Federal philosophy of science: A deconstruction—and a reconstruction», in *NYU Journal of Law and Liberty*, 5 (2): 394-435. Reprinted in Haack, 2014a: 122-55.
- 2014a: «Credulity and circumspection: Epistemological character and the ethics of belief», in *Proceedings of the American Catholic Philosophical Association*, 88: 27-47, available at https://www.pdcnet.org/acpaprocc/content/acpaprocc_2015_0999_11_16_25.
- 2014b: «Correlation and causation: The “Bradford Hill Criteria” in epidemiological, legal, and epistemological perspective», in Haack, 2014a: 239-63.
- 2014c: *Evidence matters: Science, truth, and proof in the law*. New York: Cambridge University Press.
- 2016: «Mind the analytical gap! Tracing a fault line in *Daubert*», in *Wayne Law Review*, 61 (3): 653-90.
- 2018: «In the epistemological spotlight: Epidemiological evidence in toxic torts», in *Memoria XV Congreso Panameño de Derecho Procesal*: 361-78.
- Hal, N., 2016, September 1: «Fla. high court weighs changes to expert witness standard», in *Law360*: <https://www.law360.com/articles/790136/fla-high-court-weighs-changes-to-expert-witness-standard>.
- Haug, C., 2015, December 17: «Peer-Review fraud—hacking the scientific publication process», in *New England Journal of Medicine*, 373 (25): 2393-95.
- Hennekens, C. H., et al., 1996: «Self-Reported breast implants and connective tissue diseases in female health professionals: A retrospective cohort study», in *Journal of the American Medical Association*, 275 (8): 616–21.
- Hill, A.B., 1965: «The environment and disease: Association or causation? », in *Proceedings of the Royal Society of Medicine*, 58 (5): 295-300.
- Himes, C. F., 1893, « The scientific expert in forensic procedure», in *Journal of the Franklin Institute*, 135 (6): 407-436.
- McGinty, J. C., 2018, May 5: «Shifting standards cloud autism's progress», in *Wall Street Journal*: <https://www.wsj.com/articles/on-autism-shifting-u-s-standards-cloud-the-number-of-cases-1525446011>.
- National Research Council, 2009: *Strengthening the forensic sciences in the United States*. Washington, D.C.: National Academic Press.
- Wakefield, A., et al., 1998, February 19: «Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children», in *The Lancet*, 351 (9103): 637-41.

Cases and Statutes Cited

- 2013 Fla. Laws 107 (codified as Fla. Stat. § 90.702 (2018)).
- Barefoot v. Estelle, 463 U.S. 880 (1983).
- Berry v. City of Detroit, 25 F.3d 1342 (6th Cir. 1994).
- Blum v. Merrell Dow Pharm., Inc., 33 Phila. Cnty. Rep. 193 (1996).
- Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579 (1993).
- Daubert v. Merrell Dow Pharm., Inc., 43 F.3d 1311 (9th Cir. 1995).
- Delisle v. Crane Co., 258 So. 3d 1219 (2018).
- Ferguson v. Hubbell, 97 N.Y. 507, 514 (1884).
- Frye v. United States, 293 F. 1013 (D.C. Cir. 1923).
- Gen. Elec. Co. v. Joiner, 522 U.S. 136 (1997).
- In re: Amendments to the Florida Evidence Code, No.SC19-107 (Fla. May 23, 2019).

Kaminski v. State, 63 So.2d 339 (1952).
Kumho Tire v. Carmichael, 526 U.S. 137 (1999).
Manko v. United States, 636 F. Supp. 1419 (W.D. Mo. 1986).
N.Y. & Erie R.R. Co., 62 U.S. 88, 101 (1858).
Ramirez v. State, 810 So. 2d 836 (Fla. 2001).
United States v. Bonds, 12 F.3d 540 (6th Cir. 1993).