—10—

Correlation and Causation: The "Bradford Hill Criteria" in Epidemiological, Legal, and Epistemological Perspective

> "So the litigation is used to shock the market?"
>
> "Yes, and, of course, to compensate the victims. I don't want tumors in my bladder, benign or malignant. Most jurors would feel the same way. Here's the scenario: You put together a group of fifty or so plaintiffs, and file a big lawsuit on behalf of all Dyloft patients. At precisely the same time you launch a series of television ads soliciting more cases. You hit fast and hard, and you'll get thousands of cases. The ads run coast to coast—quickie ads that'll scare folks and make them dial your toll-free number right here in D.C., where you have a warehouse full of paralegals answering the phones and doing the grunt work. It's gonna cost you some money, but if you get, say, five thousand cases, and you settle them for twenty thousand bucks each, that's one hundred million dollars. Your cut is one third."
>
> "That's outrageous!"
>
> "No, . . . that's mass tort litigation at its finest. . . ."
>
> —John Grisham[1]

1 THE BETTER PART OF VALOR

Of course, this cynical conversation (from John Grisham's novel *The King of Torts*) is fiction, not fact; still, it comes close enough to reality to bring an old proverb to mind. "Discretion is the better part of valor," I reminded myself when, shortly after cheerfully accepting an invitation to speak at a workshop on "proof of causation in mass torts," I realized I'd bitten off more than I could chew. For the fact is that mass torts—where large numbers of plaintiffs allege the same or closely similar injuries caused by the same defendant or group of

[1] John Grisham, *The King of Torts* (New York: Doubleday/Dell, 2003), 152.

defendants—raise far too many issues, legal, historical, and philosophical, for me to handle in one short paper (or even, I suspect, in one short lifetime!).

Very briefly, then: mass tort claims may involve allegations of injuries sustained when many people are involved in a large-scale accident, or are exposed to the same environmental or occupational toxin, or suffer the side effects of the same drug; or they may involve allegations of quite different kinds of injury, e.g. when many customers of car retailers or mortgage lenders, etc., or many employees of a large firm, are subject to discriminatory treatment. In the US, such claims have prompted a variety of legal responses. As the title of a 1991 article, "From 'Cases' to 'Litigation,'" suggests, aggregation of civil cases, with many plaintiffs coming together in preparation for trial or other types of adjudication, or for settlement, is a relatively recent phenomenon, arising only in the 1960s.[2] The procedures for handling such aggregated litigation include consolidation,[3] multi-district litigation,[4] bellwether trials,[5] and (the target of Grisham's cynicism) class-action lawsuits.

Each of these has its own history and its own complexities. For example, as Justice Souter explained in *Ortiz v. Fibreboard Corporation* (1999), "[a]lthough

[2] Judith Resnick, "From 'Cases' to 'Litigation,'" *Law & Contemporary Problems*, 5 (1991): 6–68, 25. As Resnick observes, now-familiar phrases like "asbestos litigation,' "the Dupont fire litigation," and "Agent Orange Litigation" indicate the shift that took place in these decades. *Ibid*. (So too, of course, do the many citations in this paper of the form "*In re* ___ Litig.")

[3] Federal Rule of Civil Procedure 42 provides that a court may consolidate actions before it that involve common questions of law or fact. See e.g., *In re* Paoli R.R. Yard PCB Litig., 35 F.3d 717 (3d Cir. 1994). The origin of Rule 42 (which came into effect in 1938) can be traced back to an Act of 1813. Suits and Costs in the Courts of the United States Act of 1813, 3 Stat 19, 21.

[4] Multi-District Litigation (MDL), as provided in 28 USC § 1407 (2006), is "[f]ederal-court litigation in which civil actions pending in different districts and involving common fact questions are transferred to a single district for coordinated pretrial proceedings, after which [they] are returned to their original districts for trial." Bryan A Garner, ed., *Black's Law Dictionary* (St. Paul, MN: Thomson Reuters, 9th ed., 2009), 1112. This provides a way of establishing a centralized forum in which related cases are treated jointly for the purposes of coordinating pre-trial discovery. See e.g., *In re* Propulsid Prods. Liab. Litig., No. 1355, 2000 WL 35621417 (J.P.M.L. Aug. 7, 2000); *In re* Vioxx Prods. Liab. Litig., 360 F. Supp. 2d 1352 (J.P.M.L. 2005) ("*Vioxx*—§ 1407 Centralization"). Eldon E. Fallon, Jeremy T. Grabill, and Robert Pitard Wynne, "Bellwether Trials in Multi-District Litigation," *Tulane Law Review* 82 (2008): 2323–67. Such "transferee courts" have also functioned to establish a mechanism for conducting bellwether trials (note 5 below). *Id*., 2332.

[5] A "bellwether" refers, literally, to the male sheep that was "belled" (i.e., had a bell put around his neck) to lead his flock. In the legal context, it is a metaphor for a "test" or "representative" case. Fallon, Grabill, and Wynne, "Bellwether Trials in Multi-District Litigation" (note 4 above), 2324. For example, the transferee court in the Vioxx MDL (note 4 above) conducted six bellwether trials, five of which resulted in verdicts for the defendants, and one in a verdict for the plaintiff. *Id*., 2335. See also Federal Judicial Center, *Manual for Complex Litigation, Fourth* (St. Paul, MN: Thomson West, 2004), 224 (on transferee courts' role in establishing bellwether trials), 360 ff. (on test cases).

representative suits have been recognized in various forms since the earliest days of English law, … class actions as we recognize them today developed as an exception to the formal rigidity of the necessary parties rule in equity."[6] Such suits turn in part on satisfaction of the requirements for class certification under Federal Rule of Civil Procedure (FRCP) 23[7]—a rule drafted before the rise of mass tort litigation on the present scale, and designed for different purposes, but then gradually adapted to the new legal needs—now including certification of a *settlement* class as well as certification of a *litigation* class.[8] And, on top of the legal complexities, there are hard policy questions about the appropriate role of the tort system vis à vis regulatory agencies such as the Food and Drug Administration (FDA) and the Environmental Protection Agency (EPA)—and a juicy philosophical question issue about how well individual justice can be served by multi-party litigation. Perhaps, one day, I shall feel able to tackle some of these; but not today.

So, as the saying goes, I have good news and I have bad news. The bad news is that I will have little to say here specifically about the "mass" in "mass torts," and nothing to say about tort litigation involving other kinds of injuries than those allegedly caused by drugs or toxic substances. The good news is that (though my illustrations will be drawn from US cases over the last couple of decades, and I won't be able to resist commenting on some epistemological weaknesses of the *Daubert* régime under which scientific evidence is now handled federally and in many states) the arguments developed here will

6 Ortiz v. Fibreboard Corp., 527 U.S. 815, 832 (1999). In this context Justice Souter cites Raymond B. Marcin, "Searching for the Origin of Class Action," *Catholic University Law Review* 23, no. 3 (1974): 515–24, Steven C. Yeazell, *From Medieval Group Litigation to the Modern Class Action* (New Haven, CT: Yale University Press, 1987), and Geoffrey C. Hazard, Jr., John L. Gedid, and Stefan Sowle, "An Historical Analysis of the Binding Effect of Class Suits," *University of Pennsylvania Law Review* 146 (1998): 1849–1948. The necessary parties rule, Justice Souter continues, citing West v. Randall, 29 F. Cas. 718, 721 (C.C.D.R.I, 1820) (No. 17,424), required "that all persons materially interested, either as plaintiffs or defendants in the subject matter … ought to be made parties to the suit, however numerous they may be."

7 "1. The class is so numerous that joinder of all members is impracticable. 2. There are questions of law or fact common to the class. 3. The claims or defenses of the representative classes are typical of the claims or defenses of the class. 4. The representative parties will fairly and adequately protect the interests of the class." FRCP 23. On the history of the Rule, and how it was gradually adapted to serve purposes that the drafters didn't intend, see generally Resnick, "From 'Cases' to 'Litigation'" (note 2 above).

8 See e.g. William W. Schwarzer, "Settlement of Mass Tort Actions: Order out of Chaos," *Cornell Law Review* 80 (1995): 837–44, 838–39; Jack B. Weinstein, *Individual Justice in Mass Tort Litigation: The Effect of Class Actions, Consolidations, and Other Multiparty Devices* (Evanston, IL: Northwestern University Press, 1995), 26, 128ff. Examples would be *In re* Vioxx Prods. Liab. Litig., 239 F.R.D. 450 (E.D. La. 2006) ("*Vioxx*—Rule 23 Certification") (a class-action suit alleging injuries from a drug); Wal-Mart Stores, Inc. v. Dukes, 131 S. Ct. 2541 (2011) (a class-action suit alleging injuries from systematic sex discrimination by an employer).

apply to questions about proof of general causation in toxic torts quite generally, not just to questions about a specific legal system at a specific time.

My topic will be the so-called "Bradford Hill criteria"—"criteria" frequently used by causation experts, featured in more than one edition of the federal *Reference Manual on Scientific Evidence*,[9] and cited in numerous toxic-tort cases in both federal and state courts—for determining when epidemiological evidence of a positive association likely indicates a causal connection. As the word "criteria" suggests, Hill is often taken to have offered a checklist of conditions satisfaction of which is necessary and/or sufficient to conclude that evidence of a causal claim is probative or, more commonly, that it is reliable enough to be admissible. But Hill himself never suggested that the factors he listed were anything more than fallible indicia of causation; and an analysis of how these factors map onto the determinants of evidential quality reveals that, indeed—though all are relevant, and all favorable, to a claim of general causation—Hill's factors are not sufficient or even, with one exception, necessary to establish such a claim, nor appropriate as a mechanical test of evidentiary reliability.

My first step will be to look closely at Hill's now-classic lecture, "The Environment and Disease," spelling out the nine factors he proposes and his comments about how they should be used (§2). Next, I will trace the ways in which these "Bradford Hill criteria" have been invoked in a range of toxic-tort cases (§3), showing that they have sometimes been badly misunderstood, and have often been applied in ways Hill didn't envisage, and probably wouldn't have endorsed. Then it will be time to put Hill's ideas in epistemological context. What Hill offers, I will argue, is best conceived as a rough sketch-map of one part of a much larger territory: evidence potentially relevant to causal claims; a rough sketch-map, moreover, focused—not surprisingly, given Hill's professional interests—on epidemiological findings, and on when intervention, especially intervention to lower the level of suspected occupational toxins, is justified. Like the sketch-map that gets you to the post office starting from the gas station with which you're familiar, Hill's list is helpful; but when it is superimposed on a more detailed map of the whole county, it is seen to be partial and incomplete (§4).

[9] Michael D. Green, D. Michal Freedman, and Leon Gordis, "Reference Guide on Epidemiology," Federal Judicial Center, *Reference Manual on Scientific Evidence* (Washington, DC: Federal Judicial Center, 2nd ed., 2000), 333–400, 375. Michael D. Green, D. Michal Freedman, and Leon Gordis, "Reference Guide on Epidemiology," Federal Judicial Center/National Research Council, *Reference Manual on Scientific Evidence* (Washington, DC: Federal Judicial Center, 3rd ed., 2011), 549–632, 600.

And this reveals, finally, some underlying reasons why Hill's ideas have so often been misinterpreted and misapplied in legal contexts: though he himself was very clear that there can be no hard-and-fast rules for determining when epidemiological evidence indicates causation, the legal *penchant* for convenient checklists has led many to construe his list of (as he says) "viewpoints" as criteria for the reliability of causation testimony; and though he himself seems to have grasped the quasi-holistic character of the determinants of evidential quality, against the backdrop of the atomistic tendencies of US evidence law, his partial sketch-map has led many astray (§5).

2 THE "BRADFORD HILL CRITERIA" IN EPIDEMIOLOGICAL CONTEXT

Sir Austin Bradford Hill (1897–1991), a respected British medical statistician, was a leading proponent of the now-standard practice of randomized clinical trials. He is well-known for his very successful textbook, *Principles of Medical Statistics*,[10] and for his work on smoking and lung cancer;[11] and best-known—at least in US legal circles—for his presidential address to the Section of Occupational Medicine at the Royal Society of Medicine in 1965, "The Environment and Disease."[12]

Of course, correlation doesn't always or necessarily indicate causation;[13] and in this lecture Hill suggests nine factors to be considered in determining whether a statistical association in a population between exposure to some substance and incidence of some disease or disorder is indicative of causation—these being the (so-called) "Bradford Hill criteria." Five of the nine factors Hill mentions were already to be found in the US Surgeon General's

[10] Austin Bradford Hill, *Principles of Medical Statistics* (London: Lancet, Ltd., 1937; 9th ed., 1971).

[11] For a brief biographical sketch see Peter Armitage, "Austin Bradford Hill" (version 3), *StatProb: The Encyclopedia Sponsored by Statistics and Probability Societies* (n.d.), available at http://statprob.com/encyclopedia/AustinBradfordHILL.html; for an account of Hill's pioneering work on randomized clinical trials, see Peter Armitage, "Bradford Hill and the Randomized Controlled Trial," *Pharmaceutical Medicine* 6 (1992): 23–37; for Hill's work on smoking and lung cancer, see Richard Doll and Austin Bradford Hill, "Smoking and Carcinoma of the Lung: Preliminary Report," *British Medical Journal* (September 30, 1950): 739–48.

[12] Austin Bradford Hill, "The Environment and Disease: Association or Causation?" *Proceedings of the Royal Society of Medicine* 58 (1965): 295–300.

[13] A recent press report illustrates the point amusingly: after the mayor of a city in California introduced recordings of birdsong to be played along a main street, the rate of minor crimes fell about 15%, and the rate of major crimes about 6%. The mayor believed this was the soothing effect of the birdsong; but skeptics point out that over the same period crime rates also fell in other cities (*without* the birdsong!). John Letzing, "A California City Is Tweeting—Chirping, Actually—in a Big Way," *Wall Street Journal*, January 17, 2012, A1, A12.

report on smoking and lung cancer, published the previous year,[14] and cited in Hill's lecture;[15] but Hill's list is articulated in much more detail—and, of course, in much more general terms, since it is not limited to only one kind of exposure or to only one disease.[16]

Hill "liked to tell people that he was trained in neither medicine nor statistics" (his degree was in economics); and the mathematical techniques he used were generally quite simple.[17] His *forte*, besides a keen eye for study design, was a robust, critical, and articulate common sense—a robust, critical, and articulate common sense much in evidence in the famous lecture on which I shall focus here.

Hill sets the stage like this:

> [W]e see that event B is associated with environmental feature A, [e.g.] that some form of respiratory illness is associated with dust in the environment. In what circumstances can we pass from this observed *association* to a verdict of *causation*? On what basis should we do so?
>
> I have no wish, nor the skill, to embark upon a philosophical discussion of the meaning of 'causation'...
>
> Disregarding any such problem in semantics, we have this situation. Our observations reveal an association between two variables ... beyond what we would care to attribute to the play of chance. What aspects of that association should we especially consider before deciding that the most likely interpretation of it is causation?[18]

[14] US Department of Health, Education, and Welfare, *Smoking and Health*, Public Health Service Publications No. 1103 (Washington, DC, 1964), chapter 3. (And well before that, there were the "Koch-Henle Postulates"; see note 45 below.)

[15] Hill, "The Environment and Disease" (note 12 above), 300.

[16] Subsequently there have been various other indications of causality proposed, besides Hill's: e.g., for determining when viruses cause cancers (Harald Zur Hausen, "Viruses in Human Cancers," *Current Science* 81, no. 5 [2001], 523–27; Brooke T. Mossman, George Klein, and Harald zur Hausen, "Modern Criteria to Determine the Etiology of Human Carcinogens," *Seminars in Cancer Biology* 14 [2004]: 449–52; Harald Zur Hausen, "Papilloma Viruses in the Causation of Human Cancers—A Brief Historical Account," *Virology*, 384 [2008]: 260–65); for determining whether there is a causal relationship polio vaccines containing SV40 and cancer (Kathleen Stratton, Donna A. Almario, and Marie C. McCormick, eds., *Immunization Safety Review* [Washington, DC: National Academies Press, 2003]); and for determining what causes cancer more generally (International Agency for Research on Cancer, "Preamble" to "A Review of Human Carcinogens," *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans* [Lyon, France: World Health Organization, vol. 100, 2008]). All these, as well as the Hill "criteria," are cited in Gannon v. United States, 571 F. Supp. 2d 615, 623–24 (E.D. Pa. 2007). See also the "Shepard Criteria" for determining teratogenicity (Thomas H. Shepard, "'Proof' of Human Teratogenicity," *Teratology* 50 [1994]: 97–98 [letters section]).

[17] Armitage, "Austin Bradford Hill" (note 11 above), 2.

[18] Hill, "The Environment and Disease" (note 12 above), 295–6.

Of course, I don't share Hill's dismissive attitude to philosophical questions as merely problems of "semantics"; indeed, I hope to convince you that a good philosophical account of the determinants of evidential quality can help us understand the real significance of his ideas about when, and to what degree, it is reasonable to infer that a statistical correlation indicates causality. But let me begin by describing Hill's proposed indicia of causation, and the very significant caveats and qualifications he introduces as he presents them:

(1) **Strength of the association**. *A strong association, i.e., a large increase in the incidence of D among those exposed to S compared to the incidence of D in those not so exposed, is one indication that the association is causal. In this context Hill mentions the enormously increased risk of scrotal cancer in chimney sweeps—200 times that of workers not exposed to tar or mineral oils;*[19] *and the very significantly increased risk of lung cancer among smokers—9–10 times the rate in non-smokers.*[20]

However, Hill adds, even when there is a strong association between S and D, the possibility should be considered that the explanation is not that exposure to S causes D, but that some other factor causes both the exposure and the disorder—e.g., that stress causes both smoking and lung cancer;[21] and even when the increased risk is very slight, the relationship may nevertheless be causal—after all, he points out, relatively few people exposed to rat urine develop Weil's disease,[22] but the connection is causal nonetheless.

(2) **Consistency**. *That different studies, conducted by different investigators and in different places and circumstances and at different times, yield the same or closely similar results, is a second indication that the connection may be causal. In this context Hill mentions that, by 1964, 29 retrospective and 7 prospective studies had found cigarette smoking to be associated with cancer of the lung.*[23]

[19] *Id.*, 295, citing Richard Doll, "Cancer," in Leslie John Witts, ed., *Medical Surveys and Clinical Trials: Some Methods and Applications of Group Research in Medicine* (London: Oxford University Press, 2nd ed., 1964), 333–49 (the original discovery was made by Percival Pott in the 18th century).

[20] Hill, "The Environment and Disease" (note 12 above), 296.

[21] Hill himself didn't draw this conclusion at first, but only after more evidence came in. See Armitage, "Austin Bradford Hill" (note 11 above), 3.

[22] Weil's disease (leptospirosis) is a bacterial disease spread by the urine of infected animals; the symptoms are fever, headache, chills, vomiting, jaundice, anemia, and rash. Some people, apparently, are much more susceptible to the disease than others. See http://www.health.ny.gov/diseases/communicable/leptospirosis/fact_sheet.htm; http://www.btninternet.com/~ringwood.canoe/Weils.htm.

[23] Hill, "The Environment and Disease" (note 12 above), 297, citing US Department of Health, Education, and Welfare, *Smoking and Health* (note 14 above).

However, Hill adds, the fact that a different inquiry yields different results doesn't necessarily undermine the original evidence;[24] and the fact that repeated studies are unavailable, or are impossible, doesn't show that the connection *isn't* causal. After all, there was only one study of the incidence of cancers of the nose and lung in nickel refiners in Wales—a study that couldn't be repeated, because the refining process was changed shortly after it was conducted.[25]

(3) **Specificity**. *That the association is restricted to specific workers, specific sites, and a specific type of disease, is a third indication that the connection may be causal.*[26]

However, Hill adds, this factor shouldn't be over-emphasized. Those nickel refiners, he points out, had an increased risk of not one but two types of cancer; and, he adds, milk can carry a whole range of infections, including scarlet fever, diphtheria, undulant fever, dysentery, and typhoid. Indeed, Hill continues, "multi-causation" is more common than single causation; so lack of specificity doesn't necessarily mean that there is no casual connection.[27]

(4) **Temporal precedence**. *For exposure to S to be a cause of D, the exposure must precede the disease.*[28]

Hill presents this factor more categorically than the others; but he immediately adds a significant qualification: that it is not always a trivial matter to determine temporal order—to tell, for example, whether such and such a diet leads to a disease, or the early stages of the disease lead to these food preferences.[29]

(5) **Biological gradient** *(dose-response curve). That the rate of D varies as the degree of exposure to S varies is a fifth indication that the connection may be casual. In this context Hill mentions that death rates from lung cancer rise linearly with the number of cigarettes smoked daily; and such a clear dose-response curve, he continues, "admits of a simple [causal] explanation."*[30]

[24] Hill, "The Environment and Disease" (note 12 above), 297.
[25] *Ibid*, citing Austin Bradford Hill, "The Statistician in Medicine" (Alfred Watson Memorial Lecture), *Journal of the Society of Actuaries* 88, no. II (1962): 178–91. (Presumably, though Hill doesn't say this, a new study under the changed circumstances *would* have been relevant to the eighth of his factors, "experiment," explained below.)
[26] Hill, "The Environment and Disease" (note 12 above), 279.
[27] *Ibid*.
[28] *Id*., 297–98.
[29] *Ibid*.
[30] *Id*., 298.

However, he adds, even if the rate of death from lung cancer was higher among *lighter* smokers, this wouldn't necessarily mean that the relation is *not* causal; moreover, biological gradient may be hard to establish, because evidence of the degree of exposure to S may be hard to come by.

(6) **Biological plausibility**. *If the purported causal connection is biologically plausible, i.e., fits in with our current knowledge of biological mechanisms, this is another indication that the connection really is causal.*[31]

However, Hill continues, "we cannot demand" that this factor be satisfied, since what is biologically plausible at a time depends on the biological knowledge then available. For a long time, he continues, there was no biological knowledge to support the statistical evidence that a woman's contracting rubella (German measles) during pregnancy could cause birth defects in her baby; but the causal inference was reasonable nonetheless. Again, John Snow's study of the opening weeks of the 1854 London cholera epidemic was strong evidence that the disease was waterborne,[32] even though it would be another 30 years before Robert Koch discovered the role of bacteria in causing disease.[33] "In short," Hill writes, "the association we observe may be one new to science or medicine and we must not dismiss it too light-heartedly as just too odd."[34]

(7) **Coherence**. *On the other hand, Hill continues, a causal interpretation of statistical data shouldn't seriously conflict with known facts about the biology of the disease in question. It is important, for example, that a causal connection between smoking and lung cancer coheres with histopathological evidence from smokers' bronchial epithelium and with the fact that we find substances known to cause skin cancer in laboratory animals in cigarette smoke.*[35]

However, Hill adds (repeating what he had already said under the heading "plausibility"), that we don't know the biological mechanism involved doesn't mean that the connection *isn't* causal.

[31] *Ibid.*

[32] *Id.*, 296, citing John Snow, *On the Mode of Communication of Cholera* (London: John Churchill, 1855) (reprinted, with other material, in *Snow on Cholera* [Cambridge, MA: Harvard Medical Library, 1936]).

[33] For a summary account of Koch's work, see Christoph Gradmann, "Heinrich Hermann Robert Koch," *Encyclopedia of Life Sciences*, available at http://www.els.net (Wiley, 2001).

[34] Hill, "The Environment and Disease" (note 12 above) 298, 299, citing Snow, *On the Mode of Communication of Cholera* (note 32 above), and Hill, "The Statistician in Medicine" (note 25 above).

[35] Hill, "The Environment and Disease" (note 12 above), 298.

(8) **Experiment**. *Evidence that, when a suspected substance is eliminated from, or reduced in, an environment, the rate of the disease or disorder goes down, Hill says, may be "the strongest support" for the causal hypothesis.*[36]

(9) **Analogy**. *Evidence that another disease or drug causes a certain kind of disorder is some reason to think that the suspect drug or chemical causes the similar disorder with which it is statistically associated. In this context, Hill mentions that the fact that rubella and Thalidomide are known to cause birth defects makes it reasonable to accept "similar but weaker evidence" with respect to exposure to another drug, or to another viral infection, during pregnancy.*[37]

Two points Hill makes here—one quite explicitly, the other initially implicitly, but stated explicitly a few pages later—will prove significant further on: that we must allow for the possibility of two kinds of *multiple causation* (more than one cause of the same disease, and more than one disease resulting from the same cause); and that the likelihood that a correlation is causal depends in some way on whether S's causing D constitutes a good *explanation* of the observed correlation.

More immediately relevant to the persistent misunderstandings of Hill's work, nowhere in this celebrated lecture does he use the word "criterion" or "criteria"; and there is every indication that his intention was quite modest. Temporal priority seems to be the only factor Hill thinks is absolutely required; and—but for the suggestion that "experimental" evidence is the strongest—he says nothing about the relative importance of these factors, suggesting rather that weighing them is a matter of judgment. In fact, as I said earlier, he expressly *denies* that any cut-and-dried rules can be given for inferring causation:

> . . . I do not believe . . . that we can usefully lay down some hard-and-fast rules of evidence that *must* be obeyed before we accept cause and effect. None of my nine [factors] can bring indisputable evidence for or against the cause-and-effect hypothesis, and none can be required as a *sine qua non*. What they can do, with greater or less strength, is to help us make up our minds on the fundamental question—is there any other way of explaining the facts before us, is there any other answer equally, or more, likely than cause and effect?[38]

[36] *Id.*, 299.
[37] *Ibid.*
[38] *Ibid.*

Nor should I omit to mention that Hill also emphatically denies that tests of statistical significance can answer this question:

> No formal tests of significance can answer these questions. Such tests can, and should, remind us of the effects that the play of chance can create, and they will instruct us in the likely magnitude of those effects. But beyond that they contribute nothing to the 'proof' of our hypothesis. . . . [I]n the USA, . . . I am told, some editors of journals will return an article because tests of significance have not been applied. Yet there are innumerable situations in which they are totally unnecessary. . . . [T]he glitter of the *t* table diverts attention from the inadequacy of the fare.[39]

In short, Hill is best understood as offering rough-and-ready guidelines—indicia, if you like, *not* "criteria"—of causality. As I said, he prefers the word "viewpoints"; and he acknowledges:

- that the factors he lists are to be applied *only* where there is already epidemiological evidence of a positive association between S and D;
- that they are not "criteria" for inferring causation—not, at any rate, on the common understanding of "criteria" as necessary and sufficient conditions;[40]
- that mathematical tests of statistical significance, though often described as ruling out the possibility that the association is due to chance, cannot determine whether an association is causal.

Several other caveats, not mentioned specifically in Hill's lecture, are also needed. First: the nine factors are not very clearly individuated: e.g., whether the incidence of D increases as the dose of or exposure to S increases (biological gradient) and whether the incidence of D falls if S is removed from or reduced in the environment (experiment) seem to be two sides of the same coin. Biological plausibility and coherence seem to be similarly interrelated: plausibility requires the causal claim to fit in with current biological knowledge, while coherence requires that it not conflict with such knowledge; and what Hill calls "analogy" is arguably just one aspect of biological plausibility. Second, with the exception of temporality, Hill's factors all come in degrees; so that we could infer, not that an association *is* causal, but only that it is *more likely to be* causal, the more this or that factor is satisfied. And third, except

[39] *Ibid.*

[40] True—despite his claim that none of the factors is a *sine qua non*—Hill suggests that temporality *is* a necessary condition; but he makes clear that no one factor, nor any combination of some or all of the nine, is sufficient.

for the apparently categorical requirement of temporal precedence and for Hill's comment that experimental evidence is especially convincing, there is no suggestion of any simple way to determine the weight to be given to one factor vis à vis the others.

3 THE "BRADFORD HILL CRITERIA" IN LEGAL CONTEXT

No doubt because he offers an apparently simple list of factors apparently easily comprehensible not only to specialists but also to attorneys, judges, and jurors, Hill's ideas about how to assess the likelihood that a statistical association indicates a causal connection have proved very attractive not only to medical scientists, but also to attorneys needing to establish causation or to impugn the other side's causation evidence, and to judges obliged to rule on the evidentiary reliability, and hence admissibility, of causation testimony.

A standard epidemiology text, Kenneth Rothman's *Modern Epidemiology*—though marred, unfortunately, by a broad streak of Popperism[41]—presents Hill's factors quite accurately, including most of his many caveats and qualifications. Rothman quotes Hill's comment that none of his nine factors "can bring indisputable evidence for or against the causal hypothesis";[42] and observes, correctly, that actually this isn't quite right: temporal precedence really *is* necessary for causation. And he refers approvingly to Hill's comment that "[a]ll scientific work is incomplete, … liable to be upset or modified by advancing knowledge."[43] Indeed.

I find no reference to Hill in the 1994 edition of the *Reference Manual on Scientific Evidence*;[44] though the chapter on epidemiology provides a list, attributed to Koch and Henle, of seven factors, all but one of which ("alternative explanations") is also found in Hill.[45] In the 2000 and 2011 editions,

[41] Kenneth Rothman, *Modern Epidemiology* (Boston: Little, Brown, 1986), 9 ff. For a detailed critique of Popper's philosophy of science, which I believe to be completely broken-backed, see "Federal Philosophy of Science: A Deconstruction—and a Reconstruction," pp. 122–55 in this volume, 125–34, and "Just Say 'No' to Logical Negativism," in Haack, *Putting Philosophy to Work: Inquiry and Its Place in Culture* (Amherst, NY: Prometheus Books, 2008; expanded ed., 2013), 179–94.

[42] Rothman, *Modern Epidemiology* (note 41 above), 19.

[43] *Id.*, 20.

[44] Linda A. Bailey, Leon Gordis, and Michael D. Green, "Reference Guide on Epidemiology," Federal Judicial Center, *Reference Manual on Scientific Evidence* (Washington, DC: Federal Judicial Center, 1st ed., 1994), 101–80, 161. (I say that "I find" no such reference because the index to this volume is quite inadequate, and of course I haven't been able to read every word of this substantial brick of a book.)

[45] The authors give no citation. But see Alfred S. Evans, "Causation and Disease: The Koch-Henle Postulates Revisited," *Yale Journal of Biology and Medicine* 49 (1976): 175–95 for a

however, Hill is mentioned specifically, and so is at least his most important caveat. In the 2000 edition we read:

> There is no formula or algorithm that can be used to assess whether a causal inference is appropriate based on these guidelines. One or more of these [Hill] factors may be absent even when a true causal relationship exists. Similarly, the existence of some factors does not ensure that a causal relationship exists.[46]

And in the 2011 edition:

> There is no formula or algorithm that can be used to assess whether a causal inference is appropriate based on these [Hill's] guidelines.... Although the drawing of causal inferences is informed by scientific expertise, it is not a determination that is made using an objective or algorithmic methodology.[47]

Nonetheless, when Hill's ideas are cited in court, these careful formulations seem to be ignored, and we are soon entangled in a great thicket of oversimplifications, misunderstandings, and misapplications.

Some of the confusions are relatively trivial. For example, apparently Hill was known to his friends as "Tony," and he only added the "Bradford" to his name late in life;[48] so perhaps it's not altogether surprising that we find him referred to by such a variety of names: not only as "Hill,"[49] "Bradford Hill,"[50]

history of the "Koch-Henle Postulates" from Jakob Henle's book on causation, published in 1840, and developments by his student, Robert Koch, in lectures of 1884 and 1890; and K. Codell Carter, "Koch's Postulates in Relation to the Work of Jakob Henle and Edwin Krebs," *Medical History* 29 (1985): 353–74 for an argument that Krebs was a more important source than Henle.

46 Green, Freedman, and Gordis, "Reference Guide on Epidemiology," *Reference Manual on Scientific Evidence*, 2nd ed. (note 9 above), 375.

47 Green, Freedman, and Gordis, "Reference Guide on Epidemiology," *Reference Manual on Scientific Evidence*, 3rd ed. (note 9 above), 600. Very oddly, the authors cite as authority for this claim, not Hill himself, but a much more recent article, Douglas Weed, "Epidemiologic Evidence and Causal Inferences," *Hematology/Oncology Clinics of North America* 124, no. 4 (2000): 797–807.

48 Armitage, "Austin Bradford Hill" (note 11 above), 1.

49 See e.g., LeBlanc v. Chevron USA, Inc., 513 F. Supp. 2d 641, 647 (E.D. La. 2007); *In re* Trasylol Prods. Liab. Litig., No. 08-MD-01928, 2010 WL 1489734, *5, *8, *10 (S.D. Fla. Mar. 8, 2010) ("*Trasylol*—Parikh"); *In re* Trasylol Prods. Liab. Litig., No. 08-MD-01928, 2010 WL 1489730, *4, *8 & n.24 (S.D. Fla. Mar. 19, 2010) ("*Trasylol*—Derschwitz"); DePyper v. Navarro, No. 83–303467-NM, 1995 WL 788828, *24 (Mich. Cir. Ct. Nov. 27, 1995); Lofgren v. Motorola, Inc., No. CV-93–05521, 1998 WL 299925, *25, *28, *29 (Ariz. Super. Ct. 1998).

50 See e.g., *In re* Breast Implant Litig., 11 F. Supp. 2d 1217, 1243 (D. Colo. 1998); Ferguson v. Riverside Sch. Dist. No. 416, No. CS-00–0097-FVS, 2002 WL 34355958, *6 (E.D. Wash. Feb. 6, 2002); Dunn v. Sandoz Pharm. Corp., 275 F. Supp. 2d 672, 67–80 (M.D.N.C. 2003); *In re* Phenylpropanolamine (PPA), 2003 WL 22417238, *16, *20, *29 (N.J. Super. Law Div. July 21, 2003); *In re* Viagra Prods. Liab. Litig., 572 F. Supp. 2d 1071, 1080 (D. Minn. 2008)

"Bradford-Hill,"[51] "A. Bradford Hill,"[52] "Dr. Bradford Hill" and "Dr. Austin Bradford Hill,"[53] "Sir Austin Bradford Hill,"[54] "Sir Bradford Hill,"[55] and (oops) "Arthur Bradford Hill,"[56] but even—this one really takes the biscuit!—"Brad Hill."[57]

And as we saw earlier, the nine factors Hill lists aren't easily individuated; so perhaps it's not altogether surprising that in one case we find an expert described as adding a tenth factor (sensitivity) to the original nine,[58] in another case an expert who "seemed unclear whether there were 8, 9, or 10 factors,"[59] and in another again, an expert who mentions seven of the factors and comments that "[t]here were two others that Hill added later."[60]

A less trivial misunderstanding is that one of Hill's nine factors, "experiment," is sometimes misinterpreted as referring not, as he quite clearly intended, to information about the result of removing the suspect substance from, or reducing it in, the workplace or the environment, but to *in vivo* experimentation, i.e., testing of drugs or suspected toxins on animals: in In re *Joint Eastern & Southern District Asbestos Litigation*,[61] for example, where the district court seems to have made this mistake; and in In re *Asbestos Litigation*,[62] where Dr. Lemen's testimony seems to reveal the same misunderstanding.

("In re *Viagra* I"); *In re* Viagra Prods. Liab. Litig., 658 F. Supp. 2d 936, 942 (D. Minn. 2009) ("In re *Viagra* II"); *In re* Stand 'N Seal Prods. Liab. Litig., 623 F. Supp. 2d1355, 1372 & n.2, 1373 (N.D. Ga. 2009); *In re* Fosamax Prods. Liab. Litig., 645 F. Supp. 2d 164, 175, 187–88 (S.D.N.Y. 2009); *In re* Asbestos Litig. (Delaware), C.A. No. 05C-11–257 ASB, 2009 WL 10344487, *7 (Del. Super. Ct. Apr. 8, 2009); *In re* Neurontin Mktg., Sales Practices, & Prods. Liab. Litig., 612 F. Supp. 2d 116, 132, 137, 153, 158 (D. Mass. 2009); Arabie v. Citgo Petroleum Corp., 49 So. 3d 529, 540 (La. Ct. App. 2010); Lewis v. Airco, Inc., No. A-3509–08T3, 2011 WL 2731880, *24 (N.J. Super. Ct. App. Div. July 15, 2011).

51 See e.g., *In re* Breast Implant Litigation (note 50 above), 1233 n.5; Soldo v. Sandoz Pharm. Corp., No. 98–1712, 2003 WL 22005007, *10 (W.D. Pa. Jan. 16, 2002) ("*Soldo*—Bradford Hill"); Matt Dietz Co. v. Torres, 198 S.W.3d 798, 803 (Tex. Ct. App. 2006). (It seems likely that some may believe there were *two* epidemiologists concerned, Bradford and Hill.)

52 See e.g., *In re* Joint E. & S. Dist. Asbestos Litig., 52 F.3d 1124, 1128 & n.2 (2d Cir. 1995).

53 See e.g., *Arabie* (note 50 above), 540.

54 See e.g., Rains v. PPG Indus., Inc., 361 F. Supp. 2d 829, 835 n.4 (S.D. Ill. 2004); *Gannon* (note 16 above), 624); Nonnon v. City of New York, 932 N.Y.S.2d 428, 433 (App. Div. 2011). (And in *Arabie* (note 50 above), 539, we find "Sir Austin Bradford-Hill.")

55 See e.g., Chapin v. A & L Parts, Inc., 732 N.W.2d 578, 584–85 (Mich. Ct. App. 2007). See also *id*, 588 & n.1 (Meter, J., concurring).

56 Milward v. Acuity Specialty Prods. Grp., Inc., 639 F.3d 11, 17 (1st Cir. 2011). (However, n.6, on the same page, citing Hill's lecture, refers to "Austin Bradford Hill.")

57 In re *Fosamax* (note 50 above), 188 n.14 (reporting testimony of Dr. Etminan).

58 In re *Phenylpropanolamine* (note 50 above), *20 (reporting testimony of Dr. Levine).

59 In re *Fosamax* (note 50 above), 188 (reporting testimony of Dr. Etminan).

60 *Arabie* (note 50 above), 540 (reporting testimony of Dr. Levy).

61 In re *Joint E. & S. Dist. Asbestos Litig.* (note 52 above), 1129.

62 In re *Asbestos Litig.* (Delaware) (note 50 above), *7 (reporting testimony of Dr. Lemen).

We also find some experts claiming to have used the Hill "criteria" *in the absence of any epidemiological evidence showing a positive association* between the suspected toxin and the alleged injury. The testimony of Dr. Kulig, a plaintiff's expert in more than one case against the manufacturer of the anti-lactation drug Parlodel, suspected of causing post-partum stroke, is a striking example:

> Dr. Kulig states that he has identified an association between Parlodel and stroke based on the pharmacological properties of bromocriptine, epidemiology, clinical studies, case reports, and animal studies.... [He testified]: "I believe causation exists because I've applied the Bradford-Hill criteria and here's what my analysis shows.... *I've taken the extra step and applied a published, generally accepted criteria* [sic] *to the analysis*.[63]

However (as Sandoz did not fail to point out), though Dr. Kulig specifically mentions epidemiology, there were in fact *no* epidemiological studies finding a positive association–-and the Hill "criteria" kick in only where there *is* such evidence.[64]

No less striking is the testimony of Dr. Etminan in In re *Fosamax*. Fosamax was prescribed to prevent the advance of osteoporosis; but—after it was given to hundreds of thousands of patients, and not just the few thousand in the manufacturers' clinical trials—was alleged to have caused severe jaw problems in some patients. Dr. Etminan, like Dr. Kulig, had applied the Hill factors, in the absence of epidemiological evidence, to "case reports, case series, prevalence studies, and animal studies"; and even testified that they are applicable in "situations where basically, you are only left with case reports." But the court points out that Rothman's *Modern Epidemiology*—which Dr. Etminan himself had described as the "Holy Grail" of epidemiology textbooks—notes that these factors apply only when we already have evidence of a positive association.[65] The manufacturer, Merck, also comes in for criticism from the court: not, however, because they argue that the Hill factors kick in only when there is *controlled* epidemiological evidence of a *statistically significant* association[66]

63 *Dunn* (note 50 above), 677 (reporting testimony of Dr. Kulig). See also *Soldo*—Bradford Hill (note 51 above), *9–10, where Dr Kulig testified to the same effect.

64 *Dunn* (note 50 above), 678. I will add (though Sandoz's attorneys apparently didn't) that temporal precedence seems to be once again an exception, necessary whether or not we have epidemiological evidence of an association.

65 In re *Fosamax* (note 50 above), 187–88.

66 *Id*., 175–t6. Thinking that Merck might have been misled by additions made by other authors to later editions of Hill's textbook on medical statistics, I checked the latest I could find, Austin Bradford Hill and I. D. Hill, *Bradford Hill's Principles of Medical Statistics* (London: Edward Arnold, 1991); but found, on the contrary, that it includes an entire section on

(embellishments the court allows to pass without comment),[67] but because they ignore findings by, among others, the American Dental Association that oral bisphosphonates can cause ONJ (osteonecrosis of the jaw).[68]

Most important, though, are the persistent misunderstandings of the status of the so-called "Hill criteria"[69]—which in one case are even described as "Sufficiency Criteria,"[70] and are also variously characterized as "considerations,"[71] "Principles,"[72] as an "evaluation scheme,"[73] *and* as a "methodology."[74] As this last term suggests, Hill's work typically figures in toxic-tort cases in the context of the Supreme Court's ruling in *Daubert* III.[75] In screening proffered scientific testimony for reliability, the *Daubert* Court observed, courts should look, not to an expert's conclusions, but to the methodology by which he arrived at those conclusions: "[t]he focus," Justice Blackmun wrote, "must be solely on principles and methodology, not on the conclusions they generate."[76] Four years later, the *Joiner* Court would back

"statistical significance and clinical importance," arguing that the common idea that a statistically significant result is invariably clinically important, and a statistically insignificant result clinically unimportant, is thoroughly confused.

67 For what it's worth, I found no case where Hill's animadversions against over-emphasis on tests of statistical significance were taken seriously; though in In re *Viagra* I (note 50 above), 1081, the court finds that the lack of statistical significance in data underlying Dr. McGwin's testimony did not make that testimony inadmissible. However, an amicus brief urging the Supreme Court to grant *certiorari* in *Daubert*, under the lead authorship of Prof. Rothman, cites Hill's lecture in support of the claim that "the talismanic phrase 'statistically significant' creates a misleading aura of infallibility totally out relation to its actual value." Brief Amici Curiae of Professors Kenneth Rothman et al. in Support of Petitioners, Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579 (1993) (No. 92–102), 1992 WL 12006438, *4, *6.

68 In re *Fosamax* (note 50 above), 186 (citing report by the American Dental Association, itself appealing to Hill's work).

69 See, e.g., the following federal cases: In re *Breast Implant Litig.* (note 50 above), 1233 n.5; *Ferguson v. Riverside* (note 50 above), *6; *Soldo*—Bradford Hill (note 51 above), *10; *Dunn* (note 50 above), 677, 678; In re *Viagra Prods. Liab. Litig.* II (note 50 above), 946; In re *Stand 'N Seal* (note 50 above), 1372 & n.2, 1373; *In re* Fosamax (note 50 above); In re *Trasylol*—Parikh (note 49 above), *8 & n.24, *10; In re *Trasylol*—Derschwitz (note 49 above), *4. See also the following state cases: *DePyper* (note 49 above), *24; *Lofgren* (note 49 above), *25; In re *Phenylpropolamine* (note 50 above), *16; *Matt Dietz* (note 51 above), 804.

70 In re *Joint E. & S. Dist. Asbestos Litig.* (note 52 above), 1128, 1130.

71 In re *Asbestos Litig.*(Delaware) (note 50 above), *7 (reporting testimony of Dr. Lemen).

72 See e.g., *Arabie* (note 50 above), 540 (reporting testimony of Dr. Levy).

73 See e.g., *Nonnon* (note 54 above), *4 (reporting testimony of Dr. Bernard).

74 See e.g., *Gannon* (note 16 above), 623 (listing various "methodologies," including Bradford Hill's, for evaluating whether an association is causal); In re *Neurontin* (note 50 above), 132 (endorsing defendant's criticism of plaintiffs' causation testimony); *DePyper* (note 49 above), *24 (reporting testimony of Dr. Preus); *Chapin* (note 55 above), 588 & n.1 (Meter, J., concurring, reporting testimony of Dr. Lemen).

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

76 *Id.*, 595.

away from the distinction of methodology vs. conclusions, observing that "methodology and conclusions are not entirely distinct from one another."[77] But expert witnesses', and courts', preoccupation with "methodology" continued; indeed, even Dennis Carlson, the Carmichaels' proffered expert on tire failure analysis in *Kumho Tire*, had a methodology: "the visual inspection method"(!).[78] No wonder, then, that the "Bradford Hill criteria" are sometimes presented as a "methodology"—or, better yet, as a "generally accepted methodology"—for determining causation; nor that satisfaction of the Hill "criteria" is sometimes taken as a touchstone for the evidentiary reliability, and hence the admissibility, of causation testimony.

So, notwithstanding Hill's quite explicit insistence that there *can be no* hard-and-fast rules for inferring causation, and despite some sober epidemiologists' acknowledgments that Hill never intended to offer such a checklist, his factors are apparently often presented to students in epidemiology as "causal criteria";[79] and are certainly often presented in court, and taken by attorneys and judges, as a decision-procedure for determining whether general causation evidence is probative, or is reliable. This kind of misunderstanding seems to take two equal and opposite forms. Expert witnesses sometimes talk as if all that's needed to establish causation is to run their evidence quickly by the list—"strength": check; "consistency": check; …, etc.—as if applying the Hill factors were a simple, mechanical task; and a few courts seem to treat Hill's factors the same way.[80] By contrast, other courts talk as if the "Bradford Hill methodology" were an arcane algorithm, comparable to a complex, technical mathematical procedure for assessing statistical significance, in which epidemiologists require specialized training: e.g., in In re *Fosamax*, where the court complains that Dr. Etminan "has not received any formal training in the application of the Bradford Hill factors."[81]

Other courts seem to assimilate the "Bradford Hill criteria" to *legal* tests for the admissibility of expert testimony. Perhaps, when the court in In re *Stand 'N Seal* observes that Dr. Spiller's testimony satisfies FRE 702 and is consistent with the Hill "criteria," it means only to suggest that consistency with the Hill factors suffices for evidentiary reliability;[82] but when, in *Matt Dietz Co.*

[77] Gen. Elec. Co. v. Joiner, 522 U.S. 136, 146 (1997) ("*Joiner* III").

[78] Kumho Tire Co. v. Carmichael, 526 U.S. 137, 146 (1999).

[79] Carl V. Phillips and Karen J. Goodman, "The Missed Lessons of Sir Austin Bradford Hill," *Epidemiologic Perspectives & Innovations*, 1, no. 3 (October 4, 2004), available at http:/www.epi-perspectives.com/content/1/1/3, 1.

[80] See e.g., *Rains* (note 54 above), 836 ff. (running Dr. Poser's [!] testimony by Hill's factors, and concluding that it meets none of them).

[81] In re *Fosamax* (note 50 above), 188 (criticizing Dr. Etminan's testimony).

[82] In re *Stand 'N Seal* (note 50 above), 1372 (admitting Dr. Spiller's testimony).

v. Torres, the court observes that "Torres does not explain how either under the Bradford-Hill criteria or the *Robinson* factors Dr. Brautbar's testimony is reliable"[83]—*Robinson* being the case in which Texas adopted *Daubert*[84]—it seems to be treating Hill's factors and the legal test for admissibility as on a par. In one case, a more-than-usually confused expert witness even suggests that the Hill "criteria" establish causation "beyond a reasonable doubt" (!).[85] And, most important for present purposes, some courts take satisfaction of the Hill "criteria" to be *sufficient*, and others take satisfaction of these factors to be *necessary*, for general causation testimony to be reliable, and hence admissible, under *Daubert*.[86]

Sufficiency: In In re *Trasylol*, for example, Dr. Parikh's testimony is ruled admissible because he gives a Hill analysis;[87] in In re *Stand 'N Seal* Dr. Spiller's testimony is ruled admissible because it is "consistent with" Hill's factors (even though he doesn't himself apply them explicitly);[88] and in In re *Neurontin* the plaintiff's causation testimony is ruled admissible because some of Hill's factors are met.[89] We see the same thing in some state court rulings: e.g., the concurrence in a Michigan case (*Chapin*) argues that it was acceptable to admit Dr. Lemen's causation testimony because it was based on a Hill analysis;[90] and in In re *Asbestos Litigation* a Delaware appeals court rules that the lower court was correct to deny a motion to exclude Dr. Lemen's causation testimony under *Daubert*, again because that testimony was based on the reliable Hill methodology.[91] Again, in *Lewis v. Airco*, arguing that Dr. Kipen correctly applied the Hill factors, a New Jersey appeals court reverses the trial court's ruling that his causation testimony is inadmissible.[92]

Necessity: In In re *Fosamax*, for example, the court excludes Dr. Etminan's general causation testimony on the grounds that he didn't apply the Hill "criteria" correctly;[93] in In re *Breast Implant Litigation* the court excludes the

[83] *Matt Dietz* (note 51 above), 804 (describing deficiencies in Torres's causation evidence).
[84] E.I du Pont de Nemours & Co. v. Robinson, 923 S.W.2d 549 (Tex. 1995).
[85] Dr. Parikh in In re *Trasylol*—Parikh (note 49 above), *8–9.
[86] But see *Nonnon* (note 54 above), *4ff for discussion of the Hill factors in the context of a *Frye* jurisdiction.
[87] In re *Trasylol*—Parikh (note 49 above), *10.
[88] In re *Stand 'N Seal* (note 50 above), 1372, 1378.
[89] In re *Neurontin* (note 50 above), 158.
[90] *Chapin* (note 55 above), 588.
[91] In re *Asbestos Litig.* (Delaware) (note 50 above), *8.
[92] *Lewis* (note 50 above), *24.
[93] In re *Fosamax* (note 50 above), 188.

testimony of plaintiff's experts Drs. Kassan, Klapper, and Blais in part because it doesn't satisfy the Hill "criteria";[94] and in *Rains* the testimony of Drs. Poser and Sultan,[95] and in *LeBlanc* the testimony of Dr. Gardner,[96] is excluded in part for the same reason. We see the same thing in some state court rulings: e.g., in *Lofgren v. Motorola*, where an Arizona court excludes the testimony of Dr. Olshan, who, it is said, is "incapable of applying the Hill criteria,"[97] and of Dr. Miller, who "could not articulate the Hill criteria and admitted that he didn't know what the criteria were or how to apply them";[98] and in *Matt Dietz Co. v. Torres*, where, as we saw earlier, a Texas court excludes Torres's proffered expert testimony on the ground that it meets neither the Hill nor the *Robinson* standard.[99]

On this matter, though, some courts disagree—holding, in effect, that satisfaction of the Hill factors is *not* necessary for admissibility: in *Ferguson v. Riverside*, for example, the court rules that there is no reason to exclude Dr. Jennings's testimony just because the dose-response relationship Hill mentions isn't present;[100] and in In re *Viagra* I Dr. McGwin's testimony is ruled admissible even though it doesn't satisfy all the Hill factors[101]—but the next year, in In re *Viagra* II, his testimony is excluded, though the court grants that "failure to satisfy the Bradford Hill criteria does not necessarily compel exclusion of an opinion as unreliable."[102]

In short, when Hill's ideas are used in court what we encounter is not the robust, critical common sense that characterizes his own writings, but a confused, and confusing, farrago of misunderstandings and misapplications. We can better understand why, I believe, when we see how the Hill factors fit into a larger epistemological picture.

94 In re *Breast Implant Litig.* (note 50 above), 1233 n.5 (Dr. Kassan); 1236 (Dr. Klapper); 1243 (Dr. Blais).

95 *Rains* (note 54 above), 835–38.

96 *LeBlanc* (note 49 above), 646–51, 663. (This case was subsequently vacated and remanded to the district court, to reconsider its ruling on the admissibility of Dr. Gardner's testimony in light of the ATSDR Benzene Toxicology Report, which at the time of its earlier decision was in draft form only. Hill's factors are not mentioned specifically.) 275 F. App'x (5th Cir. 2008). On remand, the court denied the oil company's motion to exclude Dr. Gardner's testimony, but again doesn't mention Hill. Civil Action No. 05–5485, 2009 WL 482160 (E.D. La. 2009).

97 *Lofgren* (note 49 above), *25.

98 *Id.*, *28 (excluding Dr. Olshan's testimony), *31 (excluding Dr. Miller's testimony).

99 *Matt Dietz* (note 51 above), 804.

100 *Ferguson* (note 50 above), *6, *8.

101 In re *Viagra* I (note 50 above), 1080, 1081.

102 In re *Viagra* II (note 50 above), 946.

4 THE "BRADFORD HILL CRITERIA" IN EPISTEMOLOGICAL CONTEXT

Up to now I have been speaking, with deliberate vagueness, of "causal claims" or claims of a "causal connection" between a suspect substance and some disease or disorder. Now it's time to be a little more specific, and explain that I intended such vague forms of words to refer to propositions to the effect that, at least in some instances, exposure to S contributes to bringing about the occurrence of D. This doesn't, of course, go very far towards explaining "causal."[103] It does, however, acknowledge both the possibility that a disease might have multiple causes, and the possibility that multiple diseases might be causally related to the same substance; and it allows me to distinguish general causation ("in some instances") from specific causation ("in the present instance").

I take for granted here, as I have argued elsewhere,[104] that legal degrees of proof are to be construed, not as mathematical probabilities, but as epistemic likelihoods, i.e., as degrees of warrant of the proposition at issue by the evidence presented[105] (in common-law jurisdictions, by the admissible evidence presented). And I shall of course rely on the account of warrant I have developed elsewhere[106]—an account which, to repeat, is neither simply atomistic nor fully holistic, but might best be described as a kind of articulated holism; and which is neither formal nor "wordy," but emphatically *worldly*.[107] In particular, I take for granted that the degree to which evidence warrants a proposition depends on how supportive the evidence is with respect to the proposition in question, how secure the reasons favorable to the proposition are, independent of the proposition in question, and how comprehensive the evidence is;

[103] For a brief discussion of the relation of legal to other concepts of causation, see "Risky Business: Statistical Proof of Specific Causation," pp. 264–93 in this volume, 264–66.

[104] See especially "Legal Probabilism: An Epistemological Dissent," pp. 47–77 in this volume, 56–64.

[105] It is worth noting that, in certain cases, plaintiffs may be making a claim weaker than "S causes D"—e.g., that S accelerates or "promotes" D. See e.g. *Joiner* III (note 77 above), 139 (noting that Joiner's attorney argued that exposure to PCBs promoted his early development of small-cell lung cancer). The weaker the causal claim, the less strong the evidence needed to warrant it to a given degree.

[106] See Susan Haack, *Evidence and Inquiry* (1993; 2nd ed., Amherst, NY: Prometheus Books, 2009), chapter 4; and *Defending Science—Within Reason: Between Scientism and Cynicism* (Amherst, NY: Prometheus Books, 2003), chapter 3. The theory is summarized in "Epistemology and the Law of Evidence: Problems and Projects," pp. 1–26 in this volume, 11–16.

[107] Susan Haack, *Defending Science* (note 106 above), 52. (The word "wordy," in its present use, was suggested to me by Philip Dawid.)

and that evidence may be positive, negative, or neutral with respect to a conclusion, which will be better warranted the *more* secure the *positive* reasons are, but the *less* secure the *negative* reasons are.

Relevance is a matter of degree, which is why we describe information as "highly," "somewhat," or only "marginally" relevant to some proposition at issue. Moreover, whether and if so to what degree some piece of evidence, e, is relevant to a conclusion—e.g., whether the way a drug affects ants, or geckos, or chickens, or chimps, is relevant to whether how it would affect humans—is not a simple formal relation, but can be a matter of material fact;[108] in my example, it depends on whether, and if so in what respects and to what degree, which of these creatures are physiologically like human beings—and, of course, on how similar to the drug being tested the substance they are given is.[109]

How well (combined) evidence E supports conclusion C depends on the degree of explanatory integration of the conjunction [E & C]; and how much a specific item of evidence, E_n, contributes to supporting C depends on how much better explanatorily integrated [E & E_n & C] is than [E & C] alone would be, *without* the addition of e_n. Degree of explanatory integration, in turn, is a matter of how well evidence and conclusion fit together in an explanatory account. It is enhanced to the degree that different elements of evidence interlock with each other; which in turn depends in part on their having overlapping content (e.g., referring to the same substance and the same disease)—the more so the more narrowly specific that content is (e.g. referring to small-cell lung cancer rather than, more broadly, to lung cancer, or to doxylamine rather than, more broadly, to antihistamines).[110] Supportiveness is also enhanced if the evidence interlocks with a broader explanatory story (e.g., about the type of genetic damage caused by a class of substances).

The evidence with respect to causal claims in toxic-tort cases may be drawn from any or all of a whole range of disciplines: toxicological studies analyzing the components of a suspected substance and the known effects of these or similar kinds of stuff; *in vivo* studies of the effects of the suspected substances

[108] *Id.*, 77 (where I argue that Kuhn's claim that standards of evidential quality are paradigm-relative is a kind of epistemological illusion: specifically, a misconstrual of the fact that judgments of relevance are perspectival, i.e., depend on one's factual beliefs).

[109] See "Proving Causation: The Weight of Combined Evidence," pp. 208–38 in this volume, n.86, on the problems with Grünenthal's animal studies of Thalidomide.

[110] Doxylamine succinate was the antihistamine in Bendectin (the drug at issue in *Daubert*), which was suspected of being teratogenic, as some other antihistamines are. International Agency for Research on Cancer, "Some Thyrotropic Agents," in *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans* (Lyon, France: World Health Organization, vol. 79, 2001), 145–59, available at http://www.drugs.com/mongraph/doxylamine-succinate.html.

on animals; *in vitro* studies of the effects of the suspect substance on cells; epidemiological studies of the occurrence and distribution of the disease in question in a population; and meta-analyses drawing conclusions from multiple epidemiological studies. Ideally, it would include evidence of some biological mechanism by which S causes D. It may also include physicians' case-studies and differential diagnoses; background information about human biology, about various diseases, and about what might make some subjects especially susceptible to a disease; and—at least in the context of litigation—it will often also include information about by whom studies were funded, about whether and if so how sources of funding tend to affect the results reached, about whether and, if so, where studies have been published, and about the effectiveness of pre-publication peer review as a quality-control device.[111]

The epidemiological evidence on which Hill focuses is an important part of such a congeries of evidence; but it is obviously *only* a part. And the question he is trying to answer is also, in a sense, partial, namely: suppose that we have an epidemiological study showing an increased incidence of D among those exposed to S than among those not so exposed; then what factors bear on whether, or to what degree, it is reasonable to infer a causal connection? And with the broader epistemological understanding now in place we can begin to see how Hill's factors map—somewhat roughly and unsystematically, to be sure—onto the more complex territory of the determinants of degree of warrant.

As I noted earlier in passing, Hill's fourth factor, temporal precedence, is a horse of an entirely different color from the rest: it really is a necessary condition of causality, for one thing, and it is relevant not only in the context of epidemiological evidence, but quite generally.[112] So I will focus here on the other eight Hill factors.

Suppose that e_1 represents the results of an epidemiological study showing a positive association between S and D. How do the remaining eight Hill factors bear on how well e_1 warrants the conclusion that exposure to S causes D?

- The strength of the association (factor 1) found in the study, and its specificity (factor 3), matter because, the more the incidence of D increases

[111] For more detailed discussion of the kinds of causation evidence routinely produced in toxic-tort cases, see "Proving Causation" (note 109 above) 219–21; and on issues about peer review specifically, "Peer Review and Publication: Lessons for Lawyers," pp. 156–79 in this volume, 164–72.

[112] Though, as we now know, some effects occur much longer after their cause than we might previously have imagined; e.g., that diethystilbestrol (DES), given to pregnant women to prevent miscarriage, can cause reproductive cancers in the daughters, and sterility in the sons, born to these women—decades after they took the drug. See Sarina Schrager and Beth E. Potter, "Diethylstilbestrol Exposure," *American Family Physician* 69, no. 10 (May 15, 2004), available at http://www.aafp.org/afp2004/0515/p2393.html.

with exposure to S, and the more tightly S and D are characterized, the more alternative explanations of D—other than that exposure to S causes it—are ruled out.

- If, in addition to e_1, we also have evidence from *other* studies also finding a positive association (factor 2, consistency), we have combined evidence [e_1 & e_2 & e_3] that, for the same reason, is more supportive of C than e_1 alone would be—the more so, as Hill mentions, if the studies are from different times and places.
- If, in addition to e_1, we also have evidence of a strong dose-response relationship (factor 5), then we have combined evidence [e_1 & e_4] that, for the same reason, is more supportive of C than e_1 alone would be.
- And so, similarly, if we have additional "experimental" evidence (factor 8, the other face of the dose-response relationship), evidence of biological plausibility (factor 6), of coherence (factor 7, the other face of plausibility), or of biological analogy (factor 9).

Of course, Hill was an epidemiologist, not an epistemologist. Still, he refers explicitly to explanatoriness; and a gradational understanding of warrant is implicit in his approach. He also acknowledges that it is relevant how well an epidemiological study fits into a larger explanatory account; and when he alludes to biological knowledge we don't yet have, he even half-acknowledges the role of comprehensiveness. True, in his 1965 lecture he hasn't much to say about what makes an epidemiological study better or worse—i.e., in my terminology, about its independent security—but questions about study design and execution are key themes in his famous text on medical statistics. And, even though another study with similar results can't turn a poorly-designed or poorly-conducted study into a well-designed and well-conducted one, adding the results of other studies pointing in the same direction as the one we started with itself raises the independent security of the results of the first study.

In any case, I believe Hill was right to insist that there can't be hard-and-fast rules for inferring causation. Only one of his factors (temporal precedence) is necessary, and none is sufficient, to establish a causal claim. The most one could say is that, *if* all of Hill's factors are satisfied in some degree, *and* the epidemiological and other evidence is itself reasonably secure, *and* no important relevant evidence is missing, *then* (depending in part on the degree to which his factors are satisfied), a causal conclusion is warranted to some degree.

5 ENVOI: THE PROBLEM OF EVIDENTIARY ATOMISM

So why have Hill's ideas been—by and large and on the whole, of course—so poorly handled in the US legal system?

A significant part of the explanation seems to be simply that very few attorneys and judges, and not many expert witnesses, have so much as *glanced* at Hill's (quite modestly-sized) paper, or even at the qualifications acknowledged in Rothman's textbook and in later editions of the *Reference Manual on Scientific Evidence*. Instead, it seems, they have picked up various garbled versions of Hill's ideas third-hand, so that he appears as a kind of mythological figure.

But another part of the explanation is that Hill's focus was, of course, not on issues about legal standards of admissibility or of proof, but on whether and when intervention is justified—when we should withdraw a drug from the market or, especially, clean up a workplace to remove or reduce the level of some suspected toxin. Many of the distortions of his ideas seem to be due to judges', attorneys', and expert witnesses' efforts to adapt the Hill factors to their quite different legal purposes. In fact, as we have seen, Hill's work stands in stark contrast with the legal *penchant* for simple decision-procedures; it aligns very poorly with legal threshold requirements; and—the point I want to emphasize here—it is markedly at odds with the atomistic tendencies of US evidence law generally, and of the Supreme Court's ruling in *Daubert* specifically.

This ruling encourages courts to screen each expert witness, and sometimes each item of testimony such a witness proposes to offer, for (relevance and) reliability. To be sure, the ruling doesn't explicitly require this; but the way it speaks of courts' responsibility to screen the proffered testimony of "an expert" or "the expert"[113] suggests it. And subsequent rulings in important cases like *Paoli*,[114] and Judge Kozinski's ruling when *Daubert* came back to the Ninth Circuit on remand,[115] do precisely this. Such atomistic screening means, in effect, that any item of evidence deemed not to be sufficiently warranted to satisfy the requirement of (relevance or) evidentiary reliability will be excluded. But, as I have argued, the determinants of the quality of evidence are quasi-holistic:[116] they depend, that is, in part on how well various elements interlock (contributing to explanatory integration and hence

[113] *Daubert* III (note 75 above), 589–90 ("the subject of an expert's testimony must be 'scientific … knowledge'") (footnote omitted); 592 ("the trial judge must determine … whether the expert is proposing to testify to (1) scientific knowledge that (2) will assist the trier of fact …").

[114] Note 3 above.

[115] *Daubert v. Merrell Dow Pharm.*, Inc., 43 F.3d 1311, 1320 (9th Cir. 1995) ("*Daubert* IV") (only Dr. Palmer's testimony would meet the relevance requirement, and it flunks the reliability prong of the Supreme Court's ruling in *Daubert* 1993).

[116] See "Epistemology and the Law of Evidence" (note 106 above), 15.

supportiveness), and in part on how much of the relevant evidence is included (the comprehensiveness dimension).

So how strong a plaintiff's evidence is that S sometimes causes D, or how strong a defendant's evidence is that it does not, depends in part on how tightly the components of the whole body of their expert testimony interlock, and in part on how much of the relevant information it includes.[117] Focusing exclusively on the independent security of each piece of evidence, and not on the quality of whole congeries of evidence, stands in the way of an adequate assessment of the strength of conjoined information—exactly the kind of assessment one would *want* the law to make; and can prevent parties to a suit from making what may be a legitimate claim, positive or negative, about causation.

Ironically enough, the essential epistemological point was already made in one of the amicus briefs submitted to the Supreme Court in *Daubert*, when Prof. Rothman and his epidemiologist colleagues argued that:

> … by focusing on … what conclusions, if any, can be reached from any one study, the trial court forecloses testimony about inferences that can be drawn from the combination of results reported by many such studies, even when those studies, standing alone, might not justify such inferences.[118]

All the more reason, then—after twenty years (and at a time when the influence of *Daubert* extends well beyond the US)[119]—to emphasize how much, and to what ill effect, evidentiary atomism pulls against the quasi-holistic character of warrant implicit in Hill's work, and articulated quite explicitly in my own.

[117] See also *Milward* (note 56 above), 26 (arguing that "the sum of Dr. Smith's testimony was that a weighing of the Hill factors, including biological plausibility, supported the inference that the association between benzene exposure and APL is genuinely causal").

[118] Brief for Kenneth Rothman et al. (note 67 above), *3.

[119] See "Epistemology and the Law of Evidence" (note 106 above), n. 109–113 and accompanying text.