THE USE AND MISUSE OF BRADFORD HILL IN U.S. TORT LAW

Raymond Richard Neutra,*
Carl F. Cranor, and David Gee*****

ABSTRACT: This article originated from a conference at the Royal Society of Medicine in 2016 that celebrated the 51st Anniversary of Sir Austin Bradford Hill's Presidential Address to the Royal Society: *The Environment and Disease: Association or Causation?* Courts widely recognize that Hill's published address articulates an important methodology for assisting causal inferences from epidemiological studies. A Westlaw search of federal district and appellate court decisions citing "Hill's aspects," "Hill's considerations," or "Hill's criteria" revealed numerous illustrative cases in which Hill's address has been invoked by judges ruling on the use of epidemiology in tort cases. Through analysis of numerous cases, this article illustrates that although some judges and advocates seem to understand Hill's address, many fail to correctly understand Hill's methodology and recommendations. This article argues that an understanding of the probabilistic foundations of epidemiology reveals that Hill is well justified in his recommended examinations of epidemiological (and other) evidence and his recommendations for interpreting the results of those examinations to warrant causal inference.

CITATION: Raymond Richard Neutra, Carl F. Cranor, & David Gee, *The Use and Misuse of Bradford Hill in U.S. Tort Law*, 58 Jurimetrics J. 127–162 (2018).

More than fifty years ago, Sir Austin Bradford Hill's Presidential Address to the Royal Society of Medicine—*The Environment and Disease: Association or Causation?*—identified some examinations of epidemiological data that researchers could undertake to assist inferences about the ability of environmental

*Retired Chief, Division of Environmental and Occupational Disease Control, California Department of Public Health.

**Distinguished Professor of Philosophy and Faculty Member Environmental Toxicology Graduate Program, University of California, Riverside.

***Visiting Fellow, Institute of Environment, Health and Societies, Brunel University, U.K.

For author correspondence concerning probability theory, please contact Dr. Neutra by e-mail at raymondneutra@gmail.com; and for correspondence concerning legal issues, contact Professor Cranor by e-mail at cfcranor@ucr.edu.

agents to cause disease.[1] U.S. courts have widely recognized that Hill's published address articulates an important methodology for assisting such causal inferences.[2] We conducted a Westlaw search of largely federal district and appellate court decisions citing "Hill's aspects," "Hill's considerations," or "Hill's criteria" that identified numerous illustrative instances in which Hill's address has been invoked by judges ruling on the use of epidemiology in tort cases. However, some judges and advocates seem to have mistaken interpretations of Hill's address. This article analyzes both incorrect and correct uses of the considerations Hill proposed. In addition, it argues that understanding the probabilistic foundations of epidemiology shows that Hill is well justified in his recommended examinations of epidemiological evidence and his recommendations for interpreting the results of those examinations to warrant causal inference. Understanding Hill's recommendations and probabilistic roots provides a deeper foundation for his examinations.

I. STANDARDS OF PROOF AND CERTAINTY SHOULD BE EXPRESSED IN DEGREES

Trained as an economist and statistician, Hill had early experience in statistical documentation of obvious excess occupational risks.[3]As an economist he was inclined to consider both the costs and benefits of the severity of the hazards and the social and economic costs of feasible interventions. Actions influenced by what was at stake would obviously demand different degrees of certainty about causality. Hill made this orientation explicit at the end of his famous address:

> In occupational medicine our object is usually to take action. . . . [But] it almost inevitably leads us to introduce differential standards before we convict. Thus on relatively slight evidence we might decide to restrict the use of a drug for early morning sickness in pregnant women. If we are wrong in deducing causation from association no great harm will be done. . . . But we should need very strong evidence before we made people burn a fuel in their homes that they do not like or stop smoking the cigarettes and eating the fats and sugars that they do like. In asking for very strong evidence, I would, however, repeat emphatically that this does not require crossing every "t," and swords with every critic, before we act.[4]

In this quotation Hill is reminding us that, depending on what is at stake, we require different degrees of certainty about causality to take action. A logical implication of this realization, that Hill did not make explicit, is that it would be useful to state one's certainty as a matter of degree not only in the law (e.g., "preponderance of evidence," "clear and convincing evidence," or "beyond a

1. *See generally* Austin Bradford Hill, Professor Emeritus of Med. Statistics, Univ. of London, The Environment and Disease: Association or Causation? President's Address Before the Royal Society of Medicine (Jan. 14, 1965), *in* 58 PROC. ROYAL SOC'Y MED. 295 (1965) [hereinafter Hill].

2. *See infra* Parts III, IX, X, and XI.

3. Peter Armitage, *Obituary: Sir Austin Bradford Hill, 1897–1991*, 154 J. ROYAL STAT. SOC'Y SER. A 482–84 (1991).

4. Hill, *supra* note 1, at 300.

reasonable doubt") but also so that policy makers or individual citizens could use it in a quantitative or quasi-quantitative decision analysis. For example, one does not need a "beyond a reasonable doubt" degree of suspicion of the presence of poisonous spiders lurking in the woodpile to justify slipping on work gloves every time one retrieves faggots from it for the fireplace. Before calling in a costly fumigator company, however, a greater degree of certainty of their presence might be warranted such as observations of characteristic spider webs. Around the time of his address, such Bayesian concepts were creeping into business schools.[5]

When a business decision has a chance of success and a chance of failure, one can use educated estimates of these probabilities to see if the expected yield is such that the chance is worth taking after considering the magnitude of the costs and hoped for profits. In the mid-1970s, one of us (Neutra) helped to show how in medicine, evidence from patient histories, physical examinations, and quantitative diagnostic tests could be used to warrant estimates of the probabilities of the several suspected causes of a patient's presenting problem. Furthermore, the costs, risks, and benefits of possible treatments, when combined with these probabilities in a decision tree could suggest, on balance, the course of action with the best-expected outcome.[6] To carry out this kind of analysis, a decision analyst wants more than a ranking of possible causes of the patient's symptoms based on an "inference to the best explanation."[7]

Decision analysts want to provide an estimated probability of each diagnosis for use in a decision tree. Notice that Hill's example justifying different standards "to convict" (though really, he should have said different standards "to act") pertained not just to dollars and cents, but also to differing effects on the autonomy of the persons affected. In business decisions and public policy, a justified degree of certainty is treated as if it were a probability of a result from a particular course of action. This subjective probability is multiplied by the payoff from that result to get an "expected value." In the spider example, one would multiply the small probability of being bitten by a poisonous spider by the dollar or personal cost assigned to receiving such a bite. This multiplication would yield the expected value of grabbing faggots of wood without gloves. One would compare this expected value to the tiny cost of taking time to slip on those work gloves. If this tiny cost is less than the expected disutility of grabbing faggots of wood without gloves, one is justified in donning them.

In Hill's address, his language evokes three alternative views of the nature of the thing that results from causal inference:

5. *See generally* HOWARD RAIFFA, DECISION ANALYSIS: INTRODUCTORY LECTURES ON CHOICES UNDER UNCERTAINTY (1968).

6. *See generally* MILTON C. WEINSTEIN ET AL., CLINICAL DECISION ANALYSIS (1980).

7. Igor Douven, *Abduction*, *in* STANFORD ENCYCLOPEDIA OF PHILOSOPHY (Edward N. Zalta ed., 2017), https://plato.stanford.edu/archives/sum2017/entries/abduction/.

> 1) A yes/no verdict as in the quotation above or this sentence at the beginning of the talk: "[When] *can we pass from this observed association to a verdict of causality?*"[8]
>
> 2) An argument to the best explanation: "[W]hat [is] the most likely interpretation of . . . causation?[9]
>
> 3) Producing a justified subjective probability that causality is at work.[10]

When Hill talks about the need for "slight evidence" to warrant action in some situations and "strong evidence" to warrant action in others he is hinting at a continuum of certainty. This can be dubbed a *degree of causal certification.* We avoid using "subjective probability" because it either invokes roulette wheels on the one hand or subjective whims on the other. "Degree of certainty" can also be misleading because one can be certain about things one cannot justify. A similar point can be made about Professor Susan Haack's term "degrees of warrant," because that seems to be aimed at ranking evidentiary findings as to their ability to convince.[11] Epidemiologists are making justified bets on the likely state of the world not just ranking evidence. The phrase *degree of causal certification,* suggests that some human being has followed agreed upon procedures to arrive at a particular degree of certification of causality, much in the way that, as Stephen Toulmin pointed out, a meteorologist certifies the probability of rain tomorrow.[12]

In the context of quantitative decision analysis, this degree of causal certification could derive from a range of warranted subjective probabilities elicited from a panel of experts. In a regulatory proceeding it might take the form of a justified certifying phrase such as "we certify with virtual certainty that x can cause some degree of increased risk of y." In tort law one might say "it is more likely than not that x caused disease y in the plaintiff." Thus, we agree with Hill at the points where he treats the product of causal inference to be a graduated, not a "yes or no" variable.

Certainly, Hill was focusing on public health decisions and did not mention tort law. But as this article explains, his recommended approach has often been improperly invoked in tort law rulings on evidence of causation. In a tort proceeding, a plaintiff must convince the court that the injury suffered "more likely than not" can be generally caused by a particular chemical (general causality) and that it is "more likely than not" that the chemical caused the plaintiff's specific injury (specific causality).[13]

---

8. Hill, *supra* note 1, at 295 (emphasis added).

9. *Id.* at 295, 298.

10. *Id.* at 300.

11. SUSAN HAACK, *Epistemology and the Law of Evidence: Problems and Projects*, *in* EVIDENCE MATTERS: SCIENCE, PROOF AND TRUTH IN THE LAW 4, 4 (2014).

12. STEPHEN E. TOULMIN, THE USES OF ARGUMENT 11–12 (updated ed. 2003).

13. Milward v. Acuity Specialty Prods. Grp., 639 F.3d 11, 13 (1st Cir. 2011).

II. THE BACKGROUND OF HILL'S APPROACH

What experiences led Hill to the approach recommended in his famous address? Hill and Richard Doll had, since the late 1940s, been adding their studies to the other epidemiological studies showing higher rates of lung cancer in smokers compared to nonsmokers.[14] In his address, Hill noted the dramatic upward trend in lung cancer incidence since 1900 and commented that it was what would have been expected given the upward trend in smoking habits and the apparent lung cancer risks associated with that habit.[15] The studies comparing lung cancer incidence in smokers and nonsmokers were "coherent" with time trends in smoking habit and time trends in lung cancer incidence and mortality.[16]

Up to the time of Hill's address (1965), researchers had not yet experimentally produced lung cancers in animals. Such evidence emerged only five years after Hill's address.[17]

The statistician R.A. Fisher, himself a smoker, argued that an alternative hypotheses could explain the epidemiology—that is, a genetic proclivity for smoking combined with a tendency to develop lung cancer could explain the association between the smoking habit and the risk of lung cancer.[18] This hypothesis, however, was incoherent with the association between the temporal trends in lung cancer and in the smoking habit.[19] Surely the gene prevalence could not be increasing at such a rate.[20]

Hill used lessons from his occupational experience and the smoking controversy to suggest *examining* certain *features* of epidemiological findings in order to increase or diminish one's degree of causal certification.[21] This was useful even in the face of the kind of skimpy mechanistic and toxicological data available to Hill in 1965. Such evidence is likely to be more available today to increase one's judgment of causality in the tort law and other contexts.[22] Hill used different terms interchangeably to denote both the features and the act of examining them: "aspects," "considerations," or "viewpoints." This may have contributed to the confusion about their epistemological status. For the purposes of this article, we will discuss "features" of the epidemiological findings that ought to be examined, and these acts we will call "examinations" by analogy

14. *See* Hill, *supra* note 1, at 297.

15. *Id.* at 298.

16. *Id.*

17. *See generally* E. C. Hammond et al., *Effects of Cigarette Smoking on Dogs. II. Pulmonary Neoplasms*, 21 ARCHIVES ENVTL. HEALTH 754–68 (1970).

18. Paul D. Stolley, *When Genius Errs: R.A. Fisher and the Lung Cancer Controversy*, 133 AM. J. EPIDEMIOLOGY 419–22 (1991).

19. *See, e.g.*, *id.*; *see also* Jan P. Vandenbroucke, *Those Who Were Wrong*, 130 AM. J. EPIDEMIOLOGY 3, 3–4 (1989); Sander Greenland, Letter to the Editor, *Re: "Those Who Were Wrong"*, 132 AM. J. EPIDEMIOLOGY 585–86 (1990).

20. *See* Stolley, *supra* note 18; *see also* Vandenbroucke, *supra* note 19; Greenland *supra* note 19.

21. *See generally* Hill, *supra* note 1.

22. *See* NAT'L ACADS. OF SCIS., ENG'G & MED., USING 21ST CENTURY SCIENCE TO IMPROVE RISK-RELATED EVALUATIONS 117–38 (2017).

with a physician's physical and laboratory examinations. Hill was not specifically considering how his approach might be used in either tort law or regulatory procedures when there might be rich toxicological and mechanistic evidence available or suggestive individual case histories where, for example the disease state appeared every time exposure was imposed but reliably disappeared when exposures were removed.

III. WHAT HILL DID AND DID NOT RECOMMEND

Although Hill did not himself use this analogy, he was in effect advocating that, like a diagnosing physician, all risk assessors should carry out a systematic series of examinations of the epidemiological, toxicological and mechanistic findings available to them. This should be done with the expectation that such routine examinations can sometimes, if not always, provide observed values that can either increase or diminish a degree of causal certification. Although Hill proffered a few illustrations of results that might increase the degree causal certification, he studiously avoided the temptation to specify what observed values of any of the several examinations would routinely increase or diminish causal certification.[23] If Hill had done this, he would have created criteria out of each of his recommended examinations. We will use the term *criterion* as involving two things: (1) something to be examined and (2) a particular result of that examination that must be met.

In 1964, the year before Hill's address, the United States Surgeon General's report mistakenly labeled as "criteria" the features of strength, consistency, specificity, and coherence even though it did not specify observed results that would increase or diminish the degree of causal certification.[24] Notably, Hill never used the word *criteria* to denote his recommended examinations. He differed from the renowned nineteenth-century bacteriologist Robert Koch whose "postulates" for incriminating an infectious agent as a cause of disease were indeed initially meant as examinations, each with a required result.[25] Furthermore, all of Koch's examinations had to be conducted and all had to yield a "yes" before one could declare an infectious agent to be the cause of a disease syndrome.[26]

---

23. *See generally* Hill, *supra* note 1.

24. ADVISORY COMM. TO THE SURGEON GEN., U.S. DEP'T OF HEALTH, EDUC. & WELFARE, SMOKING AND HEALTH 20 (1994).

25. *See* K. Codell Carter, *Koch's Postulates in Relation to the Work of Jacob Henle and Edwin Klebs*, 29 MED. HIST. 353, 357 (1985).

26. *See generally id.* (discussing Koch's four criteria: (1) the microorganism must be found in abundance in all organisms suffering from the disease, but should not be found in healthy organisms; (2) the microorganism must be isolated from a diseased organism and grown in pure culture; (3) the cultured microorganism should cause disease when introduced into a healthy organism; and (4) the microorganism must be reisolated from the inoculated, diseased experimental host and identified as being identical to the original specific causative agent).

While Koch's vision was initially deterministic, and seemed to promise certainty, reality soon forced him to realize reality's probabilistic nature. For example, Professor Pettenkofer and two assistants challenged the third criterion: a cultured microorganism introduced into a healthy person should cause disease.[27] They drank a solution formed from the diarrhea of a cholera victim, which teemed with cholera vibrios.[28] None contracted the disease, yet they had vibrios in their stools.[29]

Hill had more realistic expectations for the possible yield of his nine recommended examinations.

> What I do not believe—and this has been suggested—is 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 viewpoints 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 to make up our minds on the fundamental question—is there any other way of explaining the set of facts before us, is there any other answer equally, or more, likely than cause and effect?[30]

Hill is here inviting us to always consider the likelihood of the results of each examination under the causal hypothesis and then under the noncausal hypothesis and to consider which result is more likely or fits better with the causal hypothesis.[31] This is where one could say he is urging us to use inference to the best explanation.[32] However, just because the causal explanation is somewhat better would not necessarily mean that one's degree of causal certification reached the standard of "more likely than not," much less "beyond a reasonable doubt." Hill does not provide guidance on going that extra mile.

Furthermore, he did not suggest, as did Koch, that a useful causal inference would be impossible if some of the examinations were not feasible or applicable. For Hill, a degree of certainty sufficient for action can be developed with fewer examinations. Therefore, he suggested that nine examinations might be useful *if available* and *applicable*.[33] Moreover, his list of examinations does not constitute a scientific recipe for cooking up a degree of causal certification because (a) not all ingredients must be used, (b) there is no recommended amount or weighting of the several "ingredients," and (c) there is no specification of "pass" or "fail" results from the examinations.

Of Hill's nine recommended examinations, two of these pertain to items outside of the epidemiological evidence:

---

27. EDGAR ERSKINE HUME, MAX VON PETTENKOFER: HIS THEORY OF THE ETIOLOGY OF CHOLERA 54 (1927).

28. *Id.*

29. *See id.* at 57.

30. Hill, *supra* note 1, at 299.

31. *See generally id.*

32. Douven, *supra* note 7.

33. *See id.* at 299.

1. Examination of Plausibility: Did general mechanistic, toxicological or other evidence available prior to the epidemiological investigations suggest that the agent might cause the disease in question? Did subsequent mechanistic, toxicological, or other evidence increase or diminish suspicions?[34] Importantly, at the outset of his article he suggests there may even be occasions when "the general body of medical knowledge" obviates the need to detect "relationships between sickness, injury and conditions of work" because they are so well understood.[35]

2. Examination of Analogy: To what degree does epidemiological, mechanistic, toxicological or other evidence about a similar agent increase or diminish a suspicion that this agent could cause this disease?[36]

The other seven recommended examinations focused on features often found in epidemiological evidence:[37]

3. Examination of Temporality: Does the epidemiological evidence assure us the exposure preceded the disease and that having the disease did not result in acquiring the exposure?[38] If this examination establishes that the disease preceded the exposure, causal certainty is dramatically diminished. So, this gets close to being a criterion.

4. Examination of Specificity: Is this the only disease that is linked epidemiologically to this exposure?[39] This examination was originally proposed by Koch and Henle (another bacteriologist) on the erroneous belief that each infectious agent causes one and only one disease.[40] Hill mentioned this examination with severe caveats. Much later in his updated list of characteristics to consider, Evans does not include "specificity."[41]

5. Examination of Strength of Association: By "strength," Hill explicitly directed us to examine the strength of association by comparing the disease incidence in exposed people to the incidence in nonexposed people.[42] This comparison is known as the "rate ratio."[43] This examination is about comparing signal to noise and not about policy importance. A rate ratio could be very close to 1.0 and still be of policy significance if epidemiologists could detect it. For example, environmental regulations in California are triggered when

---

34. *See id.* at 298.

35. *See id.* at 295 ("A particular, and perhaps extreme, physical environment cannot fail to be harmful; a particular chemical is known to be toxic to man and therefore suspect on the factory floor.").

36. *See id.* at 299.

37. While Hill focused on epidemiological evidence because that was his expertise, as discussed later, many of his "examinations" are more widely applicable to other evidence as well. *See infra* notes 179–186 and accompanying text.

38. *See* Hill, *supra* note 1, at 297–98.

39. *See id.* at 297.

40. *See* Carter, *supra* note 25, at 370.

41. Alfred S. Evans, *Causation and Disease: The Henle-Koch Postulates Revisited*, 49 YALE J. BIOLOGY & MED. 175, 192 tbl.3 (1976).

42. Hill, *supra* note 1, at 295–96.

43. G. Tripepi et al., *Measures of Effect: Relative Risks, Odds Ratios, Risk Difference, and "Number Needed to Treat,"* 72 KIDNEY INT'L 789, 789 (2007).

typical life-long exposures would add an extra case of disease per one hundred thousand people. If a disease had a lifetime background probability in unexposed people of 10%, the added case would increase the background rate to 10.001%, equal to a rate ratio of 1.001. No epidemiological study could detect this actionable rate ratio. Thus, the examination for strength is not about whether the association is large enough to be policy relevant, and certainly not about whether this agent doubles the risk of disease, as some judges have been tempted to assert.[44] Rather the issue is whether the size of the association in a series of studies that have controlled for known confounding and other biases is greater than what experience has shown is possible from confounding, bias, or the play of chance. The finding of a very large rate ratio is relatively more likely under a causal hypothesis than under the hypothesis that only uncontrolled confounding, bias, or the play of chance has produced the association. Accordingly, the larger the observed rate ratio in a series of studies that have ruled out known bias including confounding, the more our willingness to certify causation should be increased. But how should our degree of certainty be influenced by small rate ratios? Hill commented: "We must not be too ready to dismiss a cause-and-effect hypothesis merely on the grounds that the observed association appears to be slight. There are many occasions in medicine when this is in truth so."[45] Sometimes even small relative risks may support tort law litigation.[46]

6. Examination of Biological Gradient: If the studies had information about different levels of exposure, to what degree did the associated rates of disease display a dose-response curve?[47] Unless there is biological evidence to the contrary, one would expect a risk to move upward with dose at least to a point. Sometimes there is mechanistic prior information to suggest a dose response curve of unusual shape.[48] One's confidence would be increased if the epidemiology confirmed that. However, difficulties in specifying and measuring exposure can frustrate our ability to demonstrate a clear dose-response pattern and confounders might falsely produce an apparent dose response pattern. The following thought experiment illustrates the latter point. Suppose that increasing anxiety led to an increase in numbers of cigarettes smoked. Then there would be an apparent dose response between an anxiety score and the risk of lung cancer, yet it is the hidden variable—the number of cigarettes smoked—that produced the misleading dose response.

7. Examination of Consistency: To what degree do studies of similar and different design provide similar associations between exposure and disease?[49] A variety of factors can introduce inconsistency in results even with a true cause. Different populations can differ as to their susceptibility because the agent may

44. *See infra* notes 151–170 and accompanying text.

45. Hill, *supra* note 1, at 296.

46. *See infra* notes 158–170 and accompanying text.

47. *See* Hill, *supra* note 1, at 296.

48. For different dose-response rates for hormones see Laura N. Vandenberg et al., *Hormones and Endocrine-Disrupting Chemicals: Low-Dose Effects and Nonmonotonic Dose Responses*, 33 ENDOCRINE REVS. 378 (2012).

49. Hill, *supra* note 1, at 296.

> be potentiated by other agents whose prevalence varies between study populations or even when their background incidence varies.[50] Hill therefore warned: "[D]ifferent results of a different inquiry certainly cannot be held to refute the original evidence."[51]
>
> 8. Coherence: To what degree do statistical and epidemiological studies of disease and exposure trends with gender, time, geography, and occupations show patterns that would have been predicted by the studies linking exposure and that disease?[52] Hill described coherence in the context of the Advisory Committee to the Surgeon General's report: "Thus in the discussion of lung cancer the Committee finds its association with cigarette smoking coherent with the temporal rise that has taken place in the two variables over the last generation and with the sex difference in mortality."[53]
>
> 9. Examination of Results of Experiments and Interventions: Rarely is it ethically, technically, and economically possible to carry out epidemiological experiments to see the effect on disease rates of applying or removing a hazard. Hill gives examples of spontaneous interventions where something was removed and disease rates fell. For example, he stated: "In 1923 long before any special hazard had been recognized, certain changes in the [nickel] refinery took place. No case of cancer of the nose has been observed in any man who first entered the works after that year."[54] What are the results of such experiments or such interventions? Again Hill, unlike Koch, did not require that such evidence be present.

Importantly, Hill went out of his way to explain why he was *not* suggesting that researchers examine whether conventional statistical significance was achieved in the studies they were inspecting.

> No formal tests of significance can answer those 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 these effects. Beyond that they contribute nothing to the "proof" of our hypothesis.
>
> . . . .
>
> I wonder whether the pendulum has not swung too far—not only with the attentive pupils but even with the statisticians themselves. . . . Fortunately I believe we have not yet gone so far as our friends in the USA where, 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. . . .[55]

---

50. *See id.* (discussing J.A. Heady, *False Figuring: Statistical Method in Medicine*, 89 MED. WORLD 305 (1958)).
51. *Id.* at 296–97.
52. *See id.* at 298.
53. *Id.*
54. *Id.* at 297.
55. *Id.* at 299.

IV. HILL IS NOT THE LAST WORD ON CAUSAL INFERENCE

Susan Haack reminds us that there is no such thing as an algorithmic scientific "method" that, even when followed, guarantees a true causal conclusion.[56] This is certainly true of Hill's recommended examinations. In addition, while discussing Hill, Haack emphasized the need to add reliability of the evidence and its comprehensiveness as important considerations.[57] Recently, systematic reviews strive to be transparent about criteria they use for allowing reliable studies of sufficient power into evidence and for the methods they use to assure that they cast a sufficiently wide and comprehensive net.[58] This precludes studies from being cavalierly excluded.

Also, the desired reliability for evidence evaluation requires that biases that derive from the financial interests and ideological commitments of the investigators and editors that control the gateways to publication be considered in a way that Hill did not address.[59]

New types of evidence indicating exposure, genetic, and epigenetic effects have increased opportunities for the enrichment of causal inference with accompanying challenges.[60] Thus, Hill's approaches to causal inference need supplementation.

Nonetheless, some of Hill's recommended examinations are relevant to other scientific fields. For instance, a series of animal toxicology studies will be more supportive of human causality if there is a clear dose-response of magnitude far above the observed fluctuations in previous control series for that species or there is a large difference in relative risks between exposed and control groups, and the animal has a metabolism plausibly similar to that of humans.[61] Other examinations and methodologies will likely be required as science develops. The First Circuit Court of Appeals among others has recognized that the careful exercise of nondeductive reasoning based on considerations such as

---

56. SUSAN HAACK, *Correlation and Causation: The "Bradford Hill Criteria" in Epidemiological, Legal, and Epistemological Perspective*, *in* EVIDENCE MATTERS: SCIENCE, PROOF AND TRUTH IN THE LAW, *supra* note 11, at 239, 251 (citing Michael D. Green et al., *Reference Guide on Epidemiology*, *in* REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 549, 600 (3d ed. 2011)).

57. *Id.* at 262–63.

58. Tracey J. Woodruff & Patricia Sutton, *The Navigation Guide Systematic Review Methodology: A Rigorous and Transparent Method for Translating Environmental Health Science into Better Health Outcomes*, 122 ENVTL. HEALTH PERSP., 1007, 1012 (2014).

59. NAOMI ORESKES & ERIK M. CONWAY, MERCHANTS OF DOUBT: HOW A HANDFUL OF SCIENTISTS OBSCURED THE TRUTH ON ISSUES FROM TOBACCO SMOKE TO GLOBAL WARMING 6 (2010); *see also* David Michaels & Celeste Monforton, *Manufacturing Uncertainty: Contested Science and the Protection of the Public's Health and Environment*, 95 AM. J. PUB. HEALTH, S39, S40 (2005) (showing that tobacco companies hired PR specialists as well as scientists to publish their own journal inciting doubt on the cause-and-effect theory of disease and smoking along with other industries attempts at misleading the public about negative health effects).

60. *See* NAT'L ACADS. OF SCIS., ENG'G & MED., *supra* note 22, at 2, 23.

61. *Id.* at 26; *see also* Kristen M. Fedak et al., *Applying the Bradford Hill Criteria in the 21st Century: How Data Integration Has Changed Causal Inference in Molecular Epidemiology*, EMERGING THEMES EPIDEMIOLOGY, Sept. 2015, at 6.

those mentioned above, satisfies the U.S. courts' requirement for a reliable methodology for determining causality, if it is applied properly.[62]

V. PROBABILITY THEORY JUSTIFIES HILL'S WARNINGS ABOUT ASYMMETRIC YIELD

As mentioned above, Hill does not provide guidance on how to weigh the results of the nine examinations he recommended, nor does he specify what results of each examination should increase or diminish an expert's willingness to certify causality. But he repeatedly warns that many of his examinations will have, what we will call, an asymmetric yield; they may have potential results that can increase causal certainty substantially but yet are not capable of producing results that diminish causal certainty to the same degree.[63] Similarly it is well recognized in diagnostic medicine, that some tests are able to rule in a diagnosis but cannot rule it out.

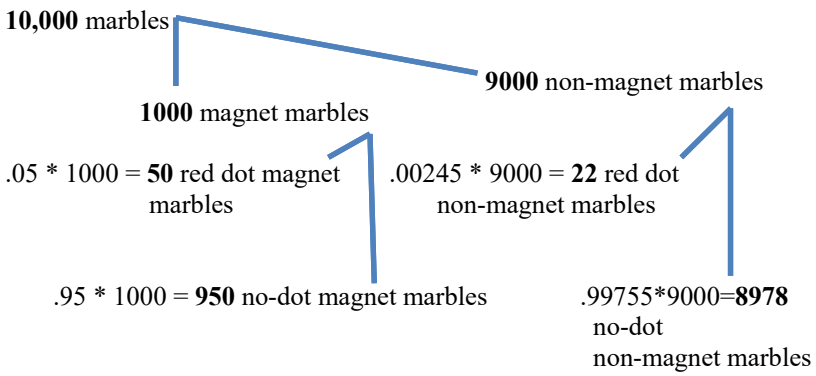
While Hill does not lay out the basis for the above mentioned asymmetry, the following example seeks to show that probabilistic principles do indeed underlie Hill's claims of both asymmetry and the importance of prior plausibility: Imagine that there are ten thousand white marbles. Ten percent of them (1000 marbles) were cast with a little magnet inside, and ninety percent (9000 marbles) were cast with no magnet inside. You are told this fact. An artisan has taken the 1000 magnet containing marbles and randomly chosen 5% of them on to which he has painted a red dot. Thus, red dots are rare among magnet containing marbles, just as strong associations are rare among causes. Then the artisan turned to the 9000 non-magnet marbles and randomly selected 0.0245% (22) of them and painted a red dot on them. So, red dots are even rarer among non-magnet marbles, just as a really strong association due to confounding, bias, or the play of chance is extremely rare in the class of non-causes. You are told these facts as well. Now he pours all the marbles in the proverbial urn evoked in elementary probability experiments and shakes up the marbles until they are thoroughly mixed. He asks you to draw one marble, and then make a bet about whether the marble will respond to the magnet he is holding. What probability (a) should you guess if he had blindfolded you so that you could not see if a red dot was present? What should you estimate (b) if the marble had no red dot, or (c) if it did have a red dot? Figure 1 provides a diagram that helps guide you to the deductively correct answer

---

62. *See* Milward v. Acuity Specialty Prods. Grp., 639 F.3d 11, 17–19 (1st Cir. 2011) ("No serious argument can be made that the weight of the evidence approach is inherently unreliable. Rather, admissibility must turn on the particular facts of the case."). For further discussion see *Schultz v. Akzo Nobel Paints,* 721 F.3d 426 (7th Cir. 2013)*,* as well as other circuits cited by the Milward court. *See Milward*, 639 F.3d at 19 ("[E]xpert testimony based on 'inference to the best explanation' may be admissible, but . . . there was no error in the district court's finding that the expert's specific theory did not have sufficient scientific support." (citing Cruz v. Bridgestone/Firestone N. Am. Tire, LLC*,* 388 F. App'x 803, 806–07 (10th Cir. 2010))); Granfield v. CSX Transp., Inc., 597 F.3d 474, 486 (1st Cir. 2010); Bitler v. A.O. Smith Corp.*,* 391 F.3d 1114, 1124 n. 5 (10th Cir. 2004).

63. *See generally* Hill, *supra* note 1.

Figure 1. Diagram to Guide Users



Here are the answers:

(a) Blind folded, you know a priori that only 10% of the marbles had magnets, so there is only this small chance that you draw one of them.

(b) Now you take off the blindfold. Suppose that after examination you see that there is no red dot on the marble. You know magnet marbles have a 95% likelihood of having no red dot and that non-magnet marbles have a 99.78% likelihood of having no red dots. Figure 1 shows the actual numbers. The probability of finding a magnet marble from among all non-red dot marbles is 950 / (950 + 8,978) = 9.6%. So, the absence of a red dot should hardly change the 10% estimate you should have given when blindfolded. This is because the two likelihoods 95% and 99.76% are similar. Indeed, their ratio is 95% / 99.76% = 0.95, pretty close to the ratio of 1.0 that would pertain if the two likelihoods were identical and thus had no discriminatory value. We can say that a relative likelihood close to 1.0 is a signal that an observed value, like the absence of a red dot, will not change one's prediction from one's a priori estimate very much. So, does that mean you should never bother to examine the marble for the presence of a red dot? Let's see.

(c) What estimate should you make if you *do* see a red dot on the marble you drew? The likelihood of finding a red dot on magnet marbles is only 5%, but the likelihood of finding one on the non-magnet marbles is much lower, that is, 0.025%. Finding a red dot conveys a relative likelihood of 5% / 0.025% = 20!

This finding should increase your estimate a lot. Look at the actual numbers in the diagram. You should be dividing 50 / (50 + 22) = 69.4%. You started with only 10% probability of having your marble respond to a magnet. Now, your best bet on the probability that you will find a magnet in that marble is well

above the tort law evidentiary standard of "more likely than not." This example shows that it is possible for a test to increase certainty a lot if it yields a positive result but yield an asymmetrically smaller impact if its result is negative.

As discussed above, this kind of asymmetry is well recognized in the field of medical diagnosis with the term rule-in-only tests.[64] When there is reliable and comprehensive evidence to generate a number for the likelihood of a result of each examination under competing hypotheses, and one has a number for the prior odds of each hypothesis being considered, one can use probability procedures to warrant a posterior probability through calculation and there will be only one possible accepted number for that posterior probability.[65] Note that the statement "the evidence speaks for itself" is false! The probative value is derived from the comparison of the known likelihoods of that observed value under the causal and then under the noncausal hypothesis.[66] This is done by calculating the relative likelihood in quantitative situations or considering what we will call the "relative fit" if one is making a qualitative argument. It is this comparison—not the naked observed finding—that can diminish or increase the updated probability beyond the prior probability.[67]

When probabilities can be assigned numbers, there is little or no room for reasonable people to disagree. As the seventeenth-century philosopher and mathematician Blaise Pascal suggested, savvy card players would do better in the long run if they apply probability theory to their game.[68]

But in the real world of causal inference, we cannot calculate posterior odds based on numerical ingredients. We both lack the necessary numbers, and the inter-correlation of our examination results would make calculations difficult even if numbers were available. In addition, there is a danger of being seduced by the apparent exactness of the results of numeric calculations. Furthermore, the desirability of transparency among other considerations may, justifiably, make researchers settle for qualitative justifications of their degree of causal certification.[69]

Nonetheless probabilistic insights and analogies still have something to offer. Decision analysts guide experts in thoughtfully eliciting a range of warranted subjective ("epistemic") probabilities after considering the relative fit of

---

64. *See* Tobias Reichlin et al., *One-Hour Rule-Out and Rule-In of Acute Myocardial Infarction Using High-Sensitivity Cardiac Troponin*, 172 ARCHIVES INTERNAL MED. 1211, 1215 (2012) (explaining that an algorithm could be used to correctly determine rule-in and rule-out percentages).

65. William C. Thompson et al., *The Role of Prior Probability in Forensic Assessments*, FRONTIERS GENETICS, Oct. 28, 2013, at 1,1.

66. *See* N. Fenton et al., *When "Neutral" Evidence Still Has Probative Value (With Implications from the Barry George Case)*, 54 SCI. & JUST. 274, 274 (2014) (describing this comparison as the "likelihood ratio").

67. *See* I.J. Good, *Weight of Evidence: A Brief Survey*, *in* BAYESIAN STATISTICS 2: PROCEEDINGS OF THE SECOND VALENCIA INTERNATIONAL MEETING, SEPTEMBER 6/10, 1983, at 249, 261 (J.M. Bernado et al. eds., 1985).

68. KEITH DEVLIN, THE UNFINISHED GAME: PASCAL, FERMAT, AND THE SEVENTEENTH-CENTURY LETTER THAT MADE THE WORLD MODERN 14–30 (2008).

69. *See* Laurence Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329, 1393 (1971).

results from various examinations and after considering the prior plausibility of causality. These elicitations are then tried out "as if" they were statistical probabilities. These probabilities inform exploratory decision analysis.[70] How certain must we be of how much disease is being caused by the status quo before we move from that status quo to cheap or expensive mitigations? Sometimes these exercises provide deeper insight to modify qualitative arguments about what we "ought" to do.[71] This kind of Bayesian sensitivity analysis has even been applied to legal evidence.[72]

Why is this insight about "rule-in" vs "rule-out" tests relevant to tort law? It is because some lawyers and their hired epidemiological consultants sometimes treat an unsupportive result of one or more of Hill's examinations as having far more evidentiary weight than is warranted.[73] Except for temporality, an unsupportive result of most Hill examinations is about equally likely under the causal and noncausal hypotheses, but lawyers and their consultants may falsely claim that such unsupportive findings should strongly diminish one's degree of causal certification to the same degree that a supportive finding would increase it.

Why are unsupportive findings so common under the causal hypothesis? The reasons vary. For example, the distribution of doses of a noxious agent delivered to a studied population is rarely high enough to produce a dramatic rate ratio even if it is of legal concern. Populations studied may vary as to susceptibility to disease or as to the presence of cofactors, and this decreases consistency in many cases. Dose response may be obscured by unavoidable difficulties in pinning down exposure in many cases. The list goes on.

Professor John Ioannidis has examined the literature and expressed doubt about the usefulness of any of Hill's recommended examinations except for consistency, temporality, and experimentation because the other examinations

---

70. *See generally* JOSEPH B. KADANE & DAVID A. SCHUM, PROBABILISTIC ANALYSIS OF THE SACCO AND VANZETTI EVIDENCE (1996) (describing experimentations with a similar probabilistic sensitivity analysis to assess the possible credibility of the pieces of legal evidence that convicted the anarchists Sacco and Vanzetti).

71. *See* Detlof von Winterfeldt et al., *Managing Potential Health Risks from Electric Powerlines: A Decision Analysis Caught in Controversy*, 24 RISK ANALYSIS 1487, 1501 (2004).

72. *See* KADANE & SCHUM, *supra* note 70.

73. One obvious example often repeated is for courts to reject epidemiological studies revealing a relative risk to exposed populations of less than 2. *See, e.g.*, Daubert v. Merrell Dow Pharm., Inc., 43 F.3d 1317 (9th Cir. 1995); Havner v. Merrell-Dow Pharm., Inc., 953 S.W.2d 706, 717–18 (1996). For a contrary view finding no authority for the "bold assertion that [standardized mortality ratios] of less than 1.5 are statistically insignificant and cannot be relied upon by a jury," see an early case, *In re Joint Eastern & Southern District Asbestos Litigation*, 52 F.3d 1124, 1134 (2d Cir. 1995). *See also infra* notes 151–170 and accompanying for a general discussion of relative risk requirements. For a more recent case endorsing a nonstatistically significant epidemiological study as relevant evidence in support of the foundation for expert testimony, see *Milward v. Acuity Specialty Products Group*, 639 F.3d 11, 23–25 (1st Cir. 2011). For a court that seems to require all of Hill's considerations, see *In re Zoloft (Sertraline Hydrochloride) Products Liability Litigation*, 26 F. Supp. 3d 446, 480–81 (E.D. Pa. 2014).

rarely increase the willingness to certify causality.[74] But the same complaint could be leveled at rule-in only diagnostic tests. Because they are rarely incriminating, these tests do not usually tip the scales of certainty, but they can be very helpful. If they are inexpensive (as are many of the Hill examinations), there is no reason not to try them out if they are applicable to the issue at hand. Nevertheless, Hill's examinations are not mandatory criteria like the Koch postulates.

VI. A PROBABILISTIC JUSTIFICATION OF THE IMPORTANCE OF PLAUSIBILITY

One of the things that Hill asks us to consider is the biological "plausibility" of the causal hypothesis.[75] This comes into play based on what was generally known before the hypothesis even came to mind. Plausibility can also come into play later when epidemiological findings stimulate a targeted examination of available information.

In probabilistic thinking, one's willingness to venture a probability on the basis of observed values is partly influenced by the likelihood ratio or the relative fit conveyed by the observed value of the findings. However, it is also influenced by the prior probability without regard to that observed evidence.[76] This can be seen by altering the above thought experiment with the magnetic marbles. Instead of 1000 / 10,000 magnetic marbles, let us have 20 / 10,000 magnetic marbles. In doing this, we have lowered the prior probability of drawing a magnetic marble dramatically. However, we will keep the respective likelihoods of red dots on the two types of marbles the same: 5% and 0.025%. Now the probability that a red dot marble is drawn from the urn would be calculated as:

$$(.05 * 20) / ((.05 * 20) + (.0025 * 9{,}980)) = 1 / (1+25) = 3.8\%$$

This result is far less than "more likely than not" although still larger than the rational guess if you were blindfolded (20 / 10,000 = 0.2%). The probative value of the observed findings as indicated by the likelihood ratio was identical in these two thought experiments but the lower a priori probability in the second example combined with the same likelihood ratio leads us to a lower degree of causality certification or posterior probability. Both the prior and relative likelihood must be considered.

How does this relate to causal inference in comprehensive toxicological and epidemiological risk assessment in tort law? We are arguing that the probability of the presence of magnets among all types of marbles is analogous to the prior plausibility that exposure x can cause disease y. The relative likelihood of a red dot observed among magnetic marbles and nonmagnetic marbles is analogous to the relative likelihood or relative fit of the collection of epidemiological and other findings under the causal and noncausal hypotheses. The prior plausibility

74. John P. A. Ioannides, *Exposure Wide Epidemiology, Revisiting Bradford Hill*, 35 STAT. MED. 1749, 1761 tbl.1 (2016).

75. Hill, *supra* note 1, at 298.

76. *See generally* RAIFFA, *supra* note 5.

should not be ignored. Once again, Hill does not explicitly reveal the probabilistic justification for his advocacy of plausibility as a thing to consider, but probability theory provides a solid backing for this advocacy.

Susan Haack (along with others) also argues that one should go beyond considering each observed value in isolation "atomistically."[77] If the observations from molecular biology, cell biology, metabolism, toxicology, and epidemiology all fit together to make a "good [causal] story" as with the last entry of a crossword puzzle, this greatly increases what she calls the evidentiary warrant.[78] While it may be rare to have a coherent good story for true causes, it is nonetheless exceedingly rare to have such a story in the class of non-causes. Thus, the good story fits the causal hypothesis far better than it fits the noncausal hypothesis and contributes something more than the observations taken in isolation.

That being said, after considering the prior plausibility and the relative fit of what we observed in many examinations, there is rarely a degree of causal certification close to either an absolute certainty of causality or absolute certainty of non-causality. Hill recognized this when he stated: *"*None of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis."[79] For the purposes of tort law, one only needs to certify to the level of "more likely than not."

VII. POTENTIAL MISUSES OF BRADFORD HILL IN U.S. TORT LAW

Hill has been erroneously invoked by judges and lawyers at two points in the legal process. First, his recommendations have sometimes been erroneously characterized to prevent scientific evidence and preclude experts who cited the evidence from being presented to the jury for consideration.[80] Second, even if experts are allowed to present their evidence and testimony to a jury, contending

77. SUSAN HAACK, *Risky Business: Statistical Proofs of Specific Causation*, *in* EVIDENCE MATTERS: SCIENCE, PROOF AND TRUTH IN THE LAW, *supra* note 11, at 287; *see also* CARL F. CRANOR, TOXIC TORTS: SCIENCE, LAW, AND THE POSSIBILITY OF JUSTICE 142–44, 150, 152, 159 (2d ed. 2016) [hereinafter TOXIC TORTS, 2d]; Michael H. Gottesman, *From* Barefoot *to* Daubert *to* Joiner*: Triple Play or Double Error?*, 40 ARIZ. L. REV. 753, 766 (1998) (showing the juxtaposition of the Supreme Court in determining unreliable expert testimony); Lisa Heinzerling, *Doubting* Daubert, 14 J.L. & POL'Y 65, 69–73 (2006); Jennifer L. Mnookin, *Atomism, Holism, and the Judicial Assessment of Evidence*, 60 UCLA L. REV. 1524, 1539 (2013); Michael S. Pardo, *Juridical Proof, Evidence, and Pragmatic Meaning: Toward Evidentiary Holism*, 95 NW. U.L. REV. 339, 399 (2000); Anthony Z. Roisman, *The Implications of* G.E. v. Joiner *for Admissibility of Expert Testimony*, 1 VT. J. ENVTL. 65, 66–67 (1999) (explaining how the Court linearly decided admissibility of expert testimony).

78. EVIDENCE MATTERS: SCIENCE, PROOF AND TRUTH IN THE LAW, *supra* note 11.

79. Hill, *supra* note 1, at 299.

80. A few courts seem to have required that all of Hill's considerations be present before an expert can testify to a jury; *see infra* notes 112–118 and accompanying text. *See also* Wade-Greaux v. Whitehall Labs., Inc., 874 F. Supp. 1441, 1450 (D.V.I. 1994). For a more recent case citing *Wade-Greaux* in which a court seemed tempted, see *In re Zoloft (Sertralinehydrochloride) Products Liability Litigation,* 176 F. Supp. 3d 449, 457 (E.D. Pa. 2014).

lawyers and their expert witnesses may misrepresent Hill's guidance to the judge and jury.

Hill's view and its statistical foundation are ignored in a series of ways, identified as E1–E6 below:

> **E1: *Demanding* unreasonable *comprehensiveness***. Types of evidence, or features, that Hill thought might be useful to examine if available or appropriate, are treated like Koch's postulates, as if no valid causal inference could be made without *all* of them.[81] Contrary to Hill's recommendation, this would be like saying that a team doctor on a climbing expedition to Mount Everest could not make a reasonable diagnosis of appendicitis unless she had a CT scanner or ultrasound device along on the trip. Of course, if the doctor were at a fancy hospital, she would be expected to use these examinations to be doubly sure, but examining the patient and the patient's story will get her well toward a diagnosis.
>
> **E2: *Construing Hill's recommended examinations as stringent "criteria."*** There has sometimes been the claim that Hill's nine examinations resemble Koch's postulates: each feature examined must display the most incriminating possible observed value, and the result considered to be exonerating if it does not reach that level.[82] In short Hill's nine examinations would be "criteria" and extremely stringent ones at that. For defense lawyers wishing to set a high bar against incriminating evidence, it is a useful ploy to misrepresent Hill this way. When epidemiologists carelessly use the phrase "Hill's criteria," they provide fodder for this ploy. This is not supported by well-respected textbooks[83] and ignores Hill's warning about the asymmetrical impact of supportive and unsupportive observed values found by the examinations.
>
> **E3: *Requiring statistical significance to 0.05***: To insist that all studies presented must reach conventional levels of statistical significance, even though this is counter to accepted textbook recommendations.[84]
>
> The existence of "null" studies, or studies with elevated risk ratios that do not attain statistical significance at the .05 level, are often erroneously considered as supportive of *no* causal relationship by attorneys and their expert witnesses, regardless of their potentially low statistical power and the difficulty of proving that no effect of policy or legal relevance exists.[85] These claims are counter to the recommendations of Hill, who specifically inveighed against

81. *See* KADANE & SCHUM, *supra* note 70.

82. *See, e.g.*, Bert Black & David E. Lilienfeld, *Epidemiologic Proof in Toxic Tort Litigation*, 52 FORDHAM L. REV. 732, 764 (1984).

83. *See generally* KENNETH J. ROTHMAN ET AL., MODERN EPIDEMIOLOGY (3d ed. 2008).

84. *Id.* at 184–88; *see also* Steven Goodman, *A Dirty Dozen: Twelve P-Value Misconceptions*, 45 SEMINARS HEMATOLOGY 135 (2008); Shinichi Nakagawa & Innes C. Cuthill, *Effect Size, Confidence Intervals and Statistical Significance: A Practical Guide for Biologists*, 82 BIOLOGICAL REV. CAMBRIDGE PHIL. SOC'Y 591 (2007).

85. *See, e.g.*, *In re* Lipitor (Atorvastatin Calcium) Mktg., Sales Practices & Prods. Liab. Litig., 174 F. Supp. 3d 911 (D.S.C. 2016); Bearden v. Honeywell Int'l, Inc., No. 3:09–CV–1035, 2015 WL 7574344, at *1 (Nov. 23, 2015); *In re* Zoloft (Sertraline Hydrochloride) Prods. Liab. Litig., 26 F. Supp. 3d 449, 455 (E.D. Pa. 2014) (explaining that an expert may not rely on nonstatistically significant studies to which to apply the BH factors).

requiring statistical significance, explicitly aimed at the best explanation, and pointed to a weight of evidence approach that considered the results of other available examinations.[86]

E4: *Requiring epidemiology evidence*. This is the demand that supportive epidemiological data must be present to make causal inferences. Under this erroneous demand, comprehensive, relevant, methodologically sound, and consistently incriminating case histories, mechanistic and animal experimental types of evidence are insufficient. Hill focused on interpreting epidemiological evidence but never claimed that it must be present to warrant a high degree of willingness to certify causality.

E5: *Requiring toxicology*. This is the claim that strong consistent epidemiology without supporting toxicology cannot lead to a high degree of causal certification.[87] In 1965 this was the situation with smoking and lung cancer and Hill's address argued that toxicological evidence was not necessary in the face of convincing epidemiological evidence.[88]

E6: *Demanding a doubling of risk*. This fallacious idea is that only when exposure in the general study population doubles the risk of disease can the degree of general and specific causation reach the tort law evidentiary standard of "more likely than not."[89] Susan Haack traces the evolution of this mistaken idea in American tort law and discusses at length why there is a logical fallacy underlying such a requirement.[90]

Without repeating all the arguments, a few examples may be given. Genetic observations might show that the plaintiff belonged to a rare subgroup that was exquisitely susceptible to an agent. In a population of such people most of the cases would indeed be due to the exposure even if this were not true for the general study population.[91] Moreover, litigants could plausibly show general causation without epidemiological data and establish specific causation with differential diagnoses or other analyses.[92] Sander Greenland and James Robins have commented at length on errors of ignoring epidemiology with relative risks less than 2.0.[93] For example, even if the cumulative number of cases in an exposed group was not doubled, an agent might accelerate the appearance of cancer, thereby losing precious person-years of healthy life for all cases.[94]

---

86. *See* Hill, *supra* note 1, at 299.

87. Wade-Greaux v. Whitehall Labs., Inc., 874 F. Supp. 1441, 1450 (D.V.I. 1994).

88. *See id.* at 298.

89. Daubert v. Merrell Dow Pharm., Inc., 43 F.3d 1317 (9th Cir. 1995).

90. HAACK, *supra* note 77, at 268–69.

91. CARL F. CRANOR, *(Almost) Equal Protection for Genetically Susceptible Subpopulations: A Hybrid Regulatory-Compensation Proposal*, *in* GENOMICS AND ENVIRONMENTAL POLICY 269 (Richard R. Sharp et al. eds., 2008) (discussing fast vs. slow acetylators).

92. *See* Zuchowicz v. United States, 140 F.3d 381, 385–87 (2d Cir. 1998); *In re* Heparin Prods. Liab. Litig., 803 F. Supp. 2d 712, 728 (N.D. Ohio 2011); TOXIC TORTS, 2d, *supra* note 77, at 257–58; *see also infra* notes 157–170 and accompanying text.

93. *See* Sander Greenland & James M. Robins, *Epidemiology, Justice, and the Probability of Causation,* 40 JURIMETRICS J. 321, 325 (2000).

94. *See, e.g.*, *id.*; *see also* Sander Greenland, *Relation of Probability of Causation to Relative Risk and Doubling Dose: A Methodological Error that Has Become a Social Problem*, 89 AM. J.

Examples of each of the erroneous interpretations of Hill's methodology are presented below. The purpose of this exposition is to help those who may be less familiar with, or have inadvertently misused Hill out of ignorance, as well as those who need to counter any misuse on the part of advocates who know better but are acting out of an excess adversarial zeal.

VIII. CHANGES IN THE LEGAL REQUIREMENTS FOR EXPERT TESTIMONY IN 1993

From 1923 to 1993 in federal and most state jurisdictions, *Frye v. United States* required judges to assess whether an expert was *qualified* to address the scientific or technical issue at the bar, and whether proposed "novel" tests or procedures were "generally accepted" by the relevant scientific community.[95] If the conditions were satisfied, judges would permit the experts to offer their *opinion* about the scientific evidence. For tests or procedures that were not novel to science and that had already been judicially recognized in a jurisdiction, there would be no judicial inquiry into the "general acceptance" by the scientific community.[96] A judge's limited role was simply to assess the qualifications of an expert and to evaluate *novel* scientific procedures to ensure that the relevant relative technical community had generally accepted them.[97]

As a result of three U.S. Supreme Court decisions, federal district court judges have been given enhanced legal duties to review the scientific foundation as well as the reliability and relevance of an expert's proposed testimony in a legal case before the expert may be permitted to testify before a jury.[98]

In *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, the Court held that the Federal Rules of Evidence[99] superseded the *Frye* rule.[100] However, judges still have a role in reviewing expert testimony and its foundation to ensure that it rests on appropriate scientific studies, and that the testimony itself is reliable and relevant to the issues in the case before an expert can testify to a jury.[101] The Court's reliability and fit inquiry was originally conceived as more liberal than

---

PUB. HEALTH 1166, 1166–69 (1998); Sander Greenland, *The Need for Critical Appraisal of Expert Witnesses in Epidemiology and Statistics*, 39 WAKE FOREST L. REV. 291, 291–310 (2004).

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

96. Donaldson v. Cent. Ill. Pub. Serv. Co., 767 N.E.2d 314, 324–27 (Ill. 2002).

97. *Frye*, 293 F. at 1014.

98. Kumho Tire Co. v. Carmichael, 526 U.S. 137, 147 (1998); Gen. Elec. Co. v. Joiner, 522 U.S. 136, 142 (1997); Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 589 (1993).

99. Federal Rule of Evidence 702. Testimony by Expert Witnesses provides:

> A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if:
>
> (a) the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;
> (b) the testimony is based on sufficient facts or data;
> (c) the testimony is the product of reliable principles and methods; and
> (d) the expert has reliably applied the principles and methods to the facts of the case.

100. *Daubert*, 509 U.S. at 587.

101. *Id.* at 589–92.

the *Frye* general acceptance requirement.[102] Only in later application and expansion did *Daubert* turn out to impose substantial new restrictions on expert testimony.[103] If an expert's proposed testimony fails on one of these dimensions, he or she may not testify. If the expert were critical to the case, the trial would be dismissed without going to the jury. These assigned responsibilities required much greater judicial engagement with and understanding of the needed science than under the *Frye* rule.[104]

In *General Electric Company v. Joiner*[105] the main appellate focus was on a procedural issue (not pertinent to our discussion). However, during its review the Court upheld the district court's atomistic or study-by-study analysis of scientific evidence.[106] According to the Federal Judicial Center's *Manual on Scientific Evidence*, this analysis poses a problem in tort cases "when plaintiff's expert relies on studies from different scientific disciplines, or studies within a discipline that present different strengths and weaknesses."[107]

Joiner appeared to lead some courts "to look at each study separately and give no consideration to those studies that cannot alone prove causation. . . . [a] slicing-and-dicing approach."[108] This method is incompatible with both Hill's approach to evidence and with how distinguished national and international scientific committees review scientific data to infer human harm. Such committees consider all the scientifically relevant data that bear on the scientific issue and then integrate human, animal, mechanistic, and other evidence to determine what hypothesis the *total* body of scientifically relevant data supports.[109]

Typically, plaintiffs are most at risk for having their experts excluded because they have the initial burden to establish sufficient evidence to change the legal status quo, which is done in part by showing their expert testimony is reliable, relevant, and appropriately supported scientifically. However, defendants occasionally have their experts excluded from testifying, which can preclude a defense to plaintiff's claim of harm.[110] Precluding defense experts appears much less frequent because defendants do not have a specific burden of proof in the trial.

With their enhanced duties to review scientific and technical testimony federal judges, typically not well prepared by legal education for these tasks, have struggled to assess the scientific support for—and the reliability and relevance of—expert testimony. This has been manifested in several ways independent of their understanding of Hill's address.

---

102. *See, e.g.*, Nimely v. City of New York, 414 F.3d 381, 396 (2d Cir. 2005).

103. *See, e.g.*, *In re* Paoli R.R. Yard PCB Litig., 35 F.3d 717, 741 (3d Cir. 1994).

104. *See generally id.* at 741–46.

105. 522 U.S. 136, 136 (1997).

106. *Id.*

107. *See* Margaret A. Berger, *The Admissibility of Expert Testimony*, *in* REFERENCE MANUAL FOR SCIENTIFIC EVIDENCE 19 (3d ed. 2011).

108. *Id.*

109. *Id.* at 20; *see also* TOXIC TORTS, 2d, *supra* note 77, at 138–41.

110. *See* Margaret Cronin Fisk, *Chicago Hope: A $28M Verdict*, NAT'L L. J., Nov. 22, 1999, at A10; *see also* Ruiz-Troche v. Pepsi Cola of P.R. Bottling Co., 161 F.3d 77, 85 (1st Cir. 1998).

IX. IDEAL EVIDENCE AND HILL'S FACTORS IN THE LAW

One concern, not specific to using epidemiological studies or Hill's considerations, is to demand what one might regard as "ideal evidence" in order to determine a cause-effect relationship between exposure and disease.[111] This might include, as it did in a few cases, requiring the plaintiffs to support their case by

> repeated, consistent epidemiological studies; . . . an animal model that duplicates the defects resulting in the human from the exposure; . . . a dose/response relationship between the exposure and the effect on the experimental fetus; and . . . the mechanism of teratogenicity of the agent should be understood and make biologic sense.[112]

While outliers, these cases suggest how courts might be tempted to demand an unreasonable degree of comprehensiveness (E1: *Demanding unreasonable comprehensiveness*).[113] They also suggest how courts might demand each of Hill's aspects to display the most incriminating observed value and if that is not achieved, the study can be disregarded (E2: Construing *Hill's recommended examinations as stringent* "*criteria*"). Courts might move in this direction by requiring mechanistic evidence in addition to epidemiological data contrary to Hill's error (E5: *Requiring Toxicology*).[114]

If such data were available, a judge might have high confidence that permitting testimony was scientifically well founded. Moreover, shortly after the *Daubert* decision some judges likely considered their task to ensure that the scientific foundation for testimony must be *correct,* as some court decisions have noted.[115] Importantly, this would be at odds with actual language of *Daubert* in which the Court recognized that some testimony should go to a jury "even though the evidence is 'shaky.'"[116] Other courts have rejected this demand for definitive evidence as inconsistent with the Federal Rules of Civil Procedure.[117]

In addition, ideal evidence requirements would likely ensure that almost no exposure-disease relationship would satisfy them because many studies would not be available. Thus, few experts would have a sufficient scientific foundation to support testimony.

---

111. CARL F. CRANOR, TRAGIC FAILURES: HOW AND WHY WE ARE HARMED BY TOXIC CHEMICALS (2017).

112. Wade-Greaux v. Whitehall Labs., Inc., 874 F. Supp. 1441, 1450 (D.V.I. 1994).

113. A recent case referencing *Wade-Greaux*, *In re* Zoloft (Sertralinehydrochloride) Prods. Liab. Litig., 176 F. Supp.3d 449, 457 (E.D. Pa. 2014), seemed similarly tempted.

114. *See* discussion of the error *supra* Part VII.

115. *See* Milward v. Acuity Specialty Prods. Grp., 639 F.3d 11, 22 (1st Cir. 2011), *rev'g* 66 F. Supp. 2d 137 (D. Mass. 2009); United States v. 14.38 Acres of Land, 80 F.3d 1074, 1078 (5th Cir. 1996); *In re* Paoli R.R. Yard PCB Litig., 35 F.3d 717, 744 (3d Cir. 1994).

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

117. *In re Paoli*, 35 F.3d at 744; *14.38 Acres of Land*, 80 F.3d at 1078; *see also Milward*, 639 F.3d at 15 (citing *Daubert*, 506 U.S. at 596).

A second judicial motivation for requiring highly certain science to support testimony likely resulted from the legacy of the *Daubert* litigation. Many epidemiological studies were available for the *Daubert* litigation. The data were also good; numerous results found no association between consumption of Bendectin during pregnancy and birth defects in children born of those pregnancies, although not all uncertainties could be removed.[118] It is plausible that judges consequently might have mistakenly believed that because there was no scientifically reliable proof of causation in the Bendectin cases without supportive epidemiology that in other proceedings when supportive epidemiology was absent causation could not be established (E4: *Requiring epidemiological evidence*). Moreover, judges not well versed about different kinds of evidence might well have considered the most direct evidence of human harm to be restricted to human studies.

The Reference Guide on Epidemiology recognizes that "most courts have appropriately declined to impose a threshold requirement that a plaintiff always must prove causation with epidemiologic evidence."[119] Furthermore, *The Admissibility of Expert Testimony,* notes that "many of the most well-respected and prestigious scientific bodies . . . consider all the relevant available scientific evidence, taken as a whole, to determine which conclusion or hypothesis regarding a causal claim is best supported by the body of evidence."[120]

Some courts have recognized this, and distinguished scientific committees often do not require epidemiological studies to infer harm to humans. For example, the International Agency for Research on Cancer (IRAC), the National Toxicology Program, and California's Proposition 65 Scientific Advisory Panel, among others, do not require epidemiological data to support findings that a substance is a probable or—in some cases—a known human carcinogen, but it is welcomed if available.[121] Thus, courts demanding that epidemiological evidence must support expert testimony are also mistaken (illustrating error E4: *Requiring epidemiological evidence*). In addition, legal requirements for good epidemiological evidence in every case would squeeze out other kinds of evidence as being scientifically and judicially inadequate for assessing causation.

Fortunately, numerous other courts have not insisted upon the presence of epidemiological evidence.[122] On this note the Tenth Circuit has stated: "We are

---

118. *See generally* MICHAEL D. GREEN, BENDECTIN AND BIRTH DEFECTS: THE CHALLENGES OF MASS TOXIC TORT LITIGATION (1996).

119. Michael D. Green et al., *Reference Guide on Epidemiology*, *in* REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 549, 610 n.183 (3d ed. 2011) (citing RESTATEMENT (THIRD) OF TORTS: LIABILITY FOR PHYSICAL AND EMOTIONAL HARM § 28 cmt. c (3) (2010)); *see also* Westberry v. Gislaved Gummi AB, 178 F.3d 257 (4th Cir. 1999); Zuchowicz v. United States, 140 F.3d 381 (2d Cir. 1998); *In re* Heparin Prods. Liab. Litig., 803 F. Supp. 2d 712 (N.D. Ohio 2011).

120. Berger, *supra* note 107, at 20.

121. TOXIC TORTS, 2d, *supra* note 77, at 307–11, 322.

122. *See, e.g.*, *Zuchowicz*, 140 F.3d at 389–90; Ambrosini v. Labarraque, 101 F.3d 129, 138–139 (D.C. Cir. 1996); Benedi v. McNeil-P.P.C., Inc., 66 F.3d 1378 (4th Cir. 1995); McCullock v. H. B. Fuller Co., 61 F.3d 1038 (2d Cir. 1995); Hopkins v. Dow Corning Corp., 33 F.3d 1116 (9th Cir. 1994); Glaser v. Thompson Med. Co., 32 F.3d 969 (6th Cir. 1994); Mendes-Silva v. United States, 980 F.2d 1482 (D.C. Cir. 1993); Kennedy v. Collagen Corp., 974 F.2d 1342 (9th Cir. 1992);

not holding that epidemiological studies are always necessary."[123] Another court, citing the Third circuit, noted that "[s]uch a requirement would 'doom from the outset all cases in which the state of research on the specific ailment or on the alleged causal agent was in its early stages.'"[124] However, do they actually practice what they preach?[125]

A related motivation for judicial certainty about testimony is that some judges appeared to overreact and demand quite stringent standards for the use of epidemiology for the foundation of testimony. Some early theorists reinforced or perhaps led to these views. Addressing causation in the law, some theorists urged strict requirements on inferences for causation from epidemiological studies drawing on the Henle-Koch-Evans postulates[126]: "Satisfaction of these [ten] criteria enables the epidemiologist to move beyond a correlation to form a biological inference that is applicable to all contemporary situations."[127] Others recognized the limitations of such stringent considerations.[128] Defendants likely reinforced such motivations, often insisting that plaintiffs must have excellent evidence. The reason: the more evidence courts legally require or the more difficult it is to satisfy evidentiary requirements, the more difficult it is for litigants to bring their cases to a jury. This motivation also exhibits errors E1: *Demanding unreasonable comprehensiveness* and E2: Construing *Hill's recommended examinations as stringent* "*criteria.*"[129]

The defense bar also often urges the need for epidemiological data in legal cases on the theory that the overall hypothesis between exposure and disease must be subject to a hypothesis test and good epidemiological studies can do that.[130] However, such studies can be quite insensitive—samples can be too small, the disease may be so rare a study will not detect it, or the time for disease latency has not elapsed. Unless studies are conducted with great attention to their design with sufficient sample sizes, sensitivity to rareness of diseases, and

---

see also Graham v. Playtex Prods., Inc., 993 F. Supp. 127, 132 (N.D.N.Y. 1998); Lakie v. SmithKline Beecham, 965 F. Supp. 49, 56 (D.D.C. 1997); *cf.* Wells v. Ortho Pharm. Corp., 788 F.2d 741, 745 (11th Cir. 1986) (reviewing a district court's decision deciding not to require the use of epidemiological evidence and instead allowing expert testimony); Globetti v. Sandoz Pharms. Corp., 111 F. Supp. 2d 1174, 1176 (N.D. Ala. 2000) (deciding not to require epidemiological evidence).

123. Norris v. Baxter Healthcare Corp*.,* 397 F.3d 878, 882 (10th Cir. 2005); *see also Benedi,* 66 F.3d at 1378; *Kennedy,* 161 F.3d at 1226; *McCullock*, 61 F.3d at 1043–44.

124. *In re* Fosamax Prods. Liab. Litig., 645 F. Supp. 2d 164, 176–77 (S.D.N.Y. 2009) (quoting Heller v. Shaw Indus., Inc., 167 F.3d 146, 155 (3d Cir. 1999)).

125. *See* Milward v. Acuity Specialty Prods. Grp., 639 F.3d 11, 24 (1st Cir. 2011) (critiquing the trial court on this point).

126. *See generally* Alfred S. Evans, *Causation and Disease: The Henle-Koch Postulates Revisited*, 49 YALE J. BIOLOGY & MED. 175 (1976).

127. Black & Lilienfeld, *supra* note 82.

128. *But see generally* Melissa M. Thompson, Comment, *Causal Inference in Epidemiology: Implications for Toxic Tort Litigation*, 71 N.C.L. REV. 247 (1992).

129. *See* discussion of the errors *supra* Part VII.

130. This point was raised by the district court in *Milward v. Acuity Specialty Products*, 664 F. Supp. 2d at 149, and rejected by the First Circuit in its appellate review, *Milward*, 639 F.3d at 23–24.

sufficient time for latency periods to run, they can miss adverse effects.[131] Granted, falsely negative studies are quite helpful to litigants seeking to protect their products, but they deny plaintiffs compensation for injuries suffered.

Finally, there may be some temptation on the part of judges and other participants to believe that excellent evidence is usually present or easily obtainable for tort cases.[132] This error is described below.

X. OTHER IDENTIFIABLE MISUNDERSTANDINGS OF HILL'S FACTORS IN THE LAW

Hill was quite clear, emphatic even, that no one of his considerations (or even all of them taken together) is a necessary condition to assist causal inferences.[133] Apart from his generic claim, *for each of his viewpoints* (except temporality) he offers a counterexample in which the consideration was not present, but causation was properly inferred nonetheless.

Judicial uses of Hill's considerations take various forms. A few courts and some commentators have argued or come close to arguing that *all* of Hill's considerations for epidemiological studies must be satisfied or an expert may testify.[134] For instance, recently a trial court held that although experts had considered the biological plausibility of Zoloft causing birth defects, "[b]ecause the experts have not adequately considered the Bradford–Hill criteria as a whole, the Court finds that their causal conclusions were not formed using reliable scientific methodology."[135] Other courts are not entirely clear, but seem to come close to this view.[136]This finding is contrary to Hill's own view and is a scientific mistake.[137]

A. Requiring Conventional and Low Statistical Significance

Both before and after the *Daubert* decision, several courts and numerous commentators have insisted that epidemiological studies must be "statistically

131. TOXIC TORTS, 2d, *supra* note 77, at 101, 209–11, 214, 225–27, 264.

132. *Id.* at 160–202.

133. Hill, *supra* note 1, at 299. Note that a cause preceding an effect *is* a necessary condition of causal inference.

134. David E. Bernstein, *The Admissibility of Scientific Evidence After* Daubert v. Merrell Dow Pharmaceuticals, Inc., 15 CARDOZO L. REV. 2139, 2181 (1994); *see* Cano v. Everest Minerals Corp., 362 F. Supp. 2d 814, 821 (W.D. Tex. 2005) (citing Merrell Dow Pharm. v. Havner, 953 S.W.2d 706, 718–19, 723–24 (Tex. 1997)) (explaining that the Hill criteria and the Henle-Koch-Evans postulates "are part of sound methodology generally accepted by the current scientific community" along with a study being statistically significant at the .05 level and "the confidence interval may not include 1.0"); *see also In re* E. and S. Dist. Asbestos Litig. 878 F. Supp. 473 (E.D.N.Y. & S.D.N.Y. 1995).

135. *See In re* Zoloft (Sertraline Hydrochloride) Prods. Liab. Litig., 26 F. Supp. 3d 446, 480–81 (E.D. Pa. 2014). This decision suggests error E1: *Demanding unreasonable comprehensiveness. See* discussion of the errors *supra* Part VII.

136. *See* Castellow v. Chevron USA, 97 F. Supp. 2d 780 (S.D. Tex. 2000); *see also In re* Breast Implant Litig., 11 F. Supp. 2d 1217 (D. Colo. 1998).

137. Hill provided counterexamples for each of his considerations, noting studies for which one or more of the considerations were missing but causation was properly inferred.

significant."[138] This requirement is typically understood to mean that the observed finding or, one more extreme, should be highly unlikely under the non-causal hypothesis with a probability < 0.05.

Some of the strictest courts have used significance testing as a "screening device" for expert testimony (committing error E3: *Requiring statistical significance to 0.05*), while other courts "are more cautious about or reject using significance testing as a necessary condition, instead recognizing that assessing the likelihood of random error is important in determining the probative value of a study . . . [and] '[t]he cold statement that a given relationship is not "statistically significant" cannot be read to mean there is no probability of a relationship.'"[139] Moreover, Hill believed that statistical significance was overrated in revealing causal relationships (and especially in the United States).[140]

Importantly, scientific studies have correctly identified causal relationships with statistical significance different from and greater than .05. For example, studies show that "ionizing radiation [has] long been known as carcinogenic in many human organ systems," yet epidemiological studies based on atomic bomb survivors [are established] "with 90 percent confidence."[141] Furthermore, *The Reference Guide on Epidemiology* notes that statistical significance at values other than 0.05 can be appropriate.[142] And Rothman and his colleagues recognize that statistical significance can be greater than 0.05, even up to 0.10.[143] Greenland has also questioned the whole logic of such tests.[144]

A variation on this concerns confidence intervals (CI). Confidence intervals provide greater information about the data by providing "the magnitude of the

---

138. *See* Brock v. Merrell Dow Pharm., Inc., 874 F.2d 307, 312 (5th Cir. 1989); *Confronting the New Challenges of Scientific Evidence*, 108 HARV. L. REV. 1481, 1541–42 (1995); *see also* Chambers v. Exxon Corp., 81 F. Supp. 2d 661, 665 (M.D. La. 2000); *In re Breast Implant*, 11 F. Supp. 2d at 1226; *Cano*, 362 F. Supp. 2d at 820 (citing *Havner*, 953 S.W.2d at 717). *See generally* Bernstein, *supra* note 134; Black & Lilienfeld, *supra* note 82.

139. Green et al., *supra* note 119, at 578 n.85 (quoting Allen v. United States, 588 F. Supp. 247, 417 (D. Utah 1984) and discussing other cases that reject the need for statistically significant studies as necessary for expert testimony including *Milward v. Acuity Specialty Products Group*, 639 F.3d 11, 24–25 (1st Cir. 2011); *Turpin v. Merrell Dow Pharmaceuticals, Inc.*, 959 F.2d 1349, 1357 (6th Cir. 1992); *DeLuca v. Merrell Dow Pharmaceuticals., Inc.*, 911 F.2d 941, 948–49 (3d Cir. 1990); *In re Viagra Products Liability Litigation*, 572 F. Supp. 2d 1071, 1090 (D. Minn. 2008); *United States v. Philip Morris USA, Inc.*, 449 F. Supp. 2d 1, 706–07 n.29 (D.D.C. 2006)).

140. Hill, *supra* note 1, at 299.

141. TOXIC TORTS, 2d, *supra* note 77, at 229 (citing H. Kato, *Cancer Mortality*, *in* CANCER IN ATOMIC BOMB SURVIVORS (I. Shigematsu & A. Kagan eds. 1986)); *see also* Julius C. McElveen, Jr. & Chris Amantea, *Legislating Risk Assessment*, 63 U. CIN. L. REV. 1553, 1556 (1995). Kato's *Cancer Mortality* was also quoted in Arthur K. Sullivan, *Classification, Pathogenesis, and Etiology of Neoplastic Diseases of the Hematopoietic System*, *in* WINTROBE'S CLINICAL HEMATOLOGY 1725, 1750 (G. Richard Lee et al. eds., 9th ed. 1993).

142. Green et al., *supra* note 119, at 577–78 (noting that although .05 is often the significance level selected, other levels can and have been used).

143. KENNETH A. ROTHMAN, MODERN EPIDEMIOLOGY 120 (1st ed. 1986); ROTHMAN ET AL., *supra* note 83, at 189.

144. Sander Greenland, *Null Misinterpretation in Statistical Testing and Its Impact on Health Risk Assessment*, 53 PREVENTIVE MED. 225, 225–28 (2011).

effect and the inherent variability in the estimate."[145] However, even confidence intervals are open to judicial mistakes. For instance one court argued, because "the CI in the McGwin *et al.* study 'includes the number 1.0,' it can be accepted that the study 'show[s] no statistically significant association between the factor and the disease.'"[146] Another court noted: "If the confidence interval includes unity, or 1.0, which means no risk, then traditional epidemiological conclusions are that the point relative risk or association found, no matter how large, is not significantly different from 1.0 or no risk."[147] Whether framed as a *P* value or as a confidence interval, the "error" is the requirement of a specified statistical significance. Thus, some courts have repeated the statistical significance error in another guise and continue to do so.[148]

International scientific committees seeking to identify toxic hazards—an exposure that *causes* an adverse health effect—reject such a view. For instance, the IARC's assessment that Roundup is a probable human carcinogen, cited two epidemiological studies that overall had an elevated relative risk (RR) above one, but the confidence interval included 1 and was considered relevant evidence: Roundup and non-Hodgkin lymphoma: RR = "1.85 (95% CI, *0.55* - 6.2)" and "1.51 (95% CI, *0.*77 - 2.94)."[149] Moreover Rothman argues that confidence intervals can asymmetrically occur around the point estimate to show the existence of an effect even though they include the number 1.0.[150]

B. Size of Effect: Requiring Epidemiological Studies with Relative Risks Greater than Two

The Ninth Circuit Court of Appeals, considering *Daubert* issues on remand, required a relative risk of two or greater in humans in support of expert testimony.[151] The court explained:

> California tort law requires plaintiffs to show not merely that Bendectin increased the likelihood of injury, but that it more likely than not caused *their* injuries. … In terms of statistical proof, this means that plaintiffs must establish not just that their mothers' ingestion of Bendectin increased somewhat the likelihood of birth defects, but that it more than doubled it—only then can it be said that Bendectin is more likely than not the source of their injury.[152]

---

145. ROTHMAN, *supra* note 143; ROTHMAN ET AL., *supra* note 83, at 189–90.

146. *In re* Viagra Prods. Liab. Litig., 572 F. Supp. 2d 1071, 1081 (D. Minn. 2008) (quoting Brock v. Merrell Dow Pharm., Inc., 874 F.2d 307, 312 (5th Cir. 1989)).

147. Smith v. Ortho Pharm. Corp., 770 F. Supp. 1561, 1576 n.49 (N.D. Ga. 1991).

148. Case Mgmt. Order No. 68, 174 F. Supp. 3d 911 (2016).

149. Int'l Agency for Research on Cancer, *Some Organophosphate Insecticides and Herbicides*, 112 IARC MONOGRAPH 342 tbl. 2.2, 395 (2017) (emphasis added).

150. ROTHMAN, *supra* note 143, at 120.

151. *Daubert*, 43 F.3d at 1321.

152. *Id.* at 1320; *see also* Cano v. Everest Minerals Corp., 362 F. Supp.2d 814, 819 (W.D. Tex. 2005). The Ninth Circuit also commits a logical error by arguing that because California law requires that plaintiffs show it is more likely than not Bendectin causes birth defects that use of epidemiological studies is the only way to show this (committing the fallacy of excluded alternatives).

A sympathetic commentator concurs, "As several courts and commentators have recognized, in the absence of special circumstances, epidemiological evidence showing a relative risk of less than 2 is ordinarily not sufficient to prove causation and avoid summary judgment."[153]

This high-profile case has been unusually influential. Numerous courts, perhaps impressed with the myriad of good epidemiological studies that were available in the *Daubert* litigation about the relation between a pregnant mother's taking Bendectin and her child's being born with shortened limb birth defects, have tended not only to require experts to support their testimony with epidemiological studies,[154] but also any epidemiological studies must have sufficient strength of effect (Hill's first consideration) to reveal a relative risk of two or greater (error E6: *Demanding a doubling of risk*).[155]

This has some superficial logic because if an "agent causes an incidence in the exposed group that is more than twice the incidence in the unexposed group (i.e., a relative risk greater than 2.0), the probability that exposure to the agent caused a similarly situated individual's disease is greater than 50%."[156] Yet this logic can be misleading as *The Reference Guide on Epidemiology* and others have noted.[157]

There are some complications to this issue: (1) What should courts *legally* require of epidemiological studies? and (2) What good prudential approaches should plaintiffs follow?

It is indeed an error concerning evidence of causation to require a relative risk of two or greater. In legal parlance this is not needed to show general causation.[158] In fact numerous scientific bodies have found causation between exposures and effects revealed by epidemiological studies where the relative risk is less and often considerably less than 2. For example, "involuntary" or secondhand tobacco smoke causes lung cancer, yet the revealed relative risk is only 1.2.[159] Combined estrogen- progestogen contraception causes breast cancer

153. *In re* Joint E. & S. Dists., 827 F. Supp. at 1028; Marder v. G.D. Searle & Co., 630 F. Supp. 1087, 1092 (D. Md. 1986); Cook v. United States, 545 F. Supp. 306, 308 (N.D. Cal. 1982); Bernstein, *supra* note 134, at 2171 (citing DeLuca v. Merrell Dow Pharm., Inc., 911 F.2d 941, 958–59 (3d Cir. 1990)). *See generally* Black & Lilienfeld, *supra* note 82.

154. Raynor v. Merrell Pharm. Inc., 104 F.3d 1371, 1375 (D.C. Cir. 1997); Wade-Greaux v. Whitehall Labs., Inc., 874 F. Supp. 1441, 1453 (D.V.I. 1994); *see also* Daubert v. Merrell Dow Pharm., Inc., 43 F.3d 1317 (9th Cir. 1995).

155. *See, e.g.*, Milward v. Acuity Specialty Prods. Grp., 639 F.3d 11, 17 (1st Cir. 2011); *Daubert*, 43 F.3d at 1321; Brock v. Merrell Dow Pharm., Inc., 874 F.2d 307, 312 (5th Cir. 1989); Richardson v. Richardson-Merrell, 857 F.2d 823, 832 (D.C. Cir. 1988); Chambers v. Exxon Corp., 81 F. Supp. 2d 661, 664 (M.D. La. 2000); *In re* Joint E. & S. Dist. Asbestos Litig., 878 F. Supp. 473 (S.D.N.Y. 1991); Renaud v. Martin Marietta Corp., 749 F. Supp. 1545, 1554 (D. Colo. 1990).

156. Green et al., *supra* note 119, at 612.

157. *Id.* at 616 n.207; *see also* Russellyn S. Carruth & Bernard D. Goldstein, *Relative Risk Greater than Two in Proof of Causation in Toxic Tort Litigation*, 41 JURIMETRICS J. 195, 206–07 (2001); Greenland, *supra* note 94.

158. Milward v. Acuity Specialty Prods. Grp., 639 F.3d 11, 17 (1st Cir. 2011).

159. U.S. DEP'T HEALTH & HUMAN SERVS., EXECUTIVE SUMMARY TO THE HEALTH CONSEQUENCES OF INVOLUNTARY EXPOSURES TO TOBACCO SMOKE: A REPORT OF THE SURGEON GENERAL 4 (excess risk among nonsmokers married to smokers was 25 percent); *id.* at 13 ("[P]ooled

with a relative risk of 1.3 to 1.43.[160] Lifetime risk of lung cancer from radon exposure among those who never smoke is about 1.5.[161] People who eat a daily portion of processed meat have an elevated risk of cancer 1.18 times greater than those who do not eat it so often.[162] In each instance the size of the effect is small, but the causal inference is certainly true. Finally, studies of radiation caused cancers in atomic bomb survivors have found "that all malignant neoplasms taken together except leukemia have a relative risk of less than two! . . . [This includes cancers of the] stomach, other parts of the digestive system, lung, and some other sites."[163] At the same time some scholars are cautious in inferences from quite small relative risks because they could be more noise than revealing causation.[164]

Requiring that studies must reveal a relative risk of two or greater is contrary to Hill's more sensitive view and to scientific norms, constituting error E6: "[Sometimes] the observed association appears to be slight. . . . Relatively few persons harboring the meningococcus fall sick of meningococcal meningitis. Relatively few persons occupationally exposed to rat's urine contract Weil's disease."[165] And, Hill certainly did not insist that a particular relative risk must be shown. Moreover, as Rothman and Greenland note, weak associations can be causal and strong associations not causal (the relation between Down syndrome and birth rank).[166]

However, general causation inquiries do not exhaust the legal issues. Plaintiffs must also show that exposure to defendant's toxicant more likely than not *caused* plaintiff's injury. When an epidemiological study that shows a relative risk of 2.0 or greater is required, this facilitates a showing of specific causation. While true, once general causation has been established, plaintiffs might have other evidence to show specific causation. They might use differential diagnosis to show plaintiffs had no other exposures likely to cause the same disease or provide mechanistic or genetic evidence to show that plaintiff was a member of a class that was more vulnerable to the agent than the majority of people in an

---

evidence indicates a 20 to 30 percent increase in the risk of lung cancer from secondhand smoke exposure associated with living with a smoker.").

160. INT'L AGENCY FOR RESEARCH ON CANCER, COMBINED ESTROGEN−PROGESTOGEN CONTRACEPTIVES AND COMBINED ESTROGEN−PROGESTOGEN MENOPAUSAL THERAPY 54 (2007)

161. Jing Chen, *Canadian Lung Cancer Relative Risk from Radon Exposure for Short Periods in Childhood Compared to a Lifetime*, 10 INT'L J. ENVTL. RES. & PUB. HEALTH 1916, 1923 (2013).

162. Press Release, Int'l Agency for Research on Cancer, World Health Org., IARC Monographs Evaluate Consumption of Red Meat and Processed Meat (Oct. 26, 2015), https://www.iarc.fr/en/media-centre/pr/2015/pdfs/pr240_E.pdf [https://perma.cc/R53Z-AEEH] ("The experts concluded that each 50 gram portion of processed meat eaten daily increases the risk of colorectal cancer by 18%.").

163. TOXIC TORTS, 2d, *supra* note 77, at 229–30 (citing Kato, *supra* note 141); *see also* McElveen & Amantea, *supra* note 141.

164. Michael D. Green, *The Future of Proportional Liability*, *in* EXPLORING TORT LAW (Stuart Madden ed., 2005); Samuel M. Lesko & Allen A. Mitchell, *The Use of Randomized Controlled Trials for Pharmacoepidemiology Studies*, *in* PHARMACOEPIDEMIOLOGY 599, 601 (Brian L. Strom ed., 4th ed. 2005)

165. Hill, *supra* note 1, at 296.

166. ROTHMAN ET AL., *supra* note 83, at 24–25.

epidemiological study. For instance, genetic susceptibility renders subgroups of children and adults more susceptible to polycyclic aromatic hydrocarbons, typical byproducts of combustion,[167] to organophosphate pesticides,[168] and to methylmercury.[169] This is where plaintiffs' prudence enters. Because there may be some limits to differential diagnoses or availability to mechanistic or genetic data, or other approaches, for example, if there are numerous possible causes of the adverse effect and some major ones cannot be ruled out, plaintiffs may not have a sufficient foundation or sufficiently reliable testimony to support specific causation. At this point judges may have grounds for ruling against them.

Nevertheless, decisions that concern plaintiffs' prudence should not dictate judges' legal rulings (so that they always require epidemiological studies with a RR > 2). If plaintiffs cannot establish specific causation as more likely than not, they will lose. However, our point is that judges should recognize there is flexibility for showing specific causation beyond just having epidemiological studies with RR > 2.

The Reference Guide on Epidemiology concurs: "Having additional evidence that bears on individual causation has led a few courts to conclude that a plaintiff may satisfy his or her burden of production even if a relative risk less than 2.0 emerges from the epidemiologic evidence."[170]

C. Specificity

Occasionally courts appear to require specificity—noting that courts should adopt "strong skepticism when *one* agent is alleged to cause *many different* types of disease."[171] However, Hill had doubts about this factor, and numerous scientists have shown the problem with it.[172]

D. Biologically Plausible Mechanism

A few courts have been unduly optimistic that "[t]he underlying predicates of any cause-and-effect medical testimony are that medical science understands

---

167. Frederica P. Perera et al., *Molecular Epidemiologic Research on the Effects of Environmental Pollutants on the Fetus*" 107 ENVTL. HEALTH PERSP. 451 (1999).

168. Maryse F. Bouchard et al., *Prenatal Exposure to Organophosphate Pesticides and IQ in 7-Year-Old Children*, 119 ENVTL. HEALTH PERSP. 1189 (2011).

169. J. Julvez et al., *Prenatal Methylmercury Exposure and Genetic Predisposition to Cognitive Deficit at Age 8 Years*, 24 EPIDEMIOLOGY 643 (2013).

170. *In re* Hanford Nuclear Reservation Litig., 292 F.3d 1124, 1137 (9th Cir. 2002); Magistrini v. One Hour Martinizing Dry Cleaning, 180 F. Supp. 2d 584, 606 (D.N.J. 2002); Miller v. Pfizer, Inc., 196 F. Supp. 2d 1062, 1079 (D. Kan. 2002); Pafford v. Sec'y, Dep't of Health & Human Servs., 64 Fed. Cl. 19 (2005) (acknowledging that epidemiologic studies finding a relative risk of less than 2.0 can provide supporting evidence of causation); Green et al., *supra* note 119, at 616.

171. *In re* Actos (Pioglitazone) Prods. Liab. Litig., No. 12-CV-00064, 2013 WL 6825953, at *12 (W.D. La. Dec. 20, 2013).

172. ROTHMAN ET AL., *supra* note 83, at 25; Brian MacMahon & Thomas F. Pugh, *Causes and Entities of Disease*, *in* METHODS OF PREVENTIVE MEDICINE 11, 16 (D. W. Clark & B. MacMahon eds. 1967); Mervyn Susser, *Judgment and Causal Inferences Criteria in Epidemiologic Studies*, 105 AM. J. EPIDEMIOLOGY 105 (1977).

the physiological process by which a particular disease or syndrome develops and knows what factors cause the process to occur,"[173] and have required specificity. In a case concerning an injury to an animal, defendants urged specificity on a trial court as the appropriate legal principle by arguing: "the specific knowledge of the precise physiological cause of Night Passage's death is a prerequisite to admissibility."[174] However, the Sixth Circuit correctly rejected this view, finding that it is, "entirely too strict a standard when considering the admissibility of their testimony." Some other courts have also rejected this requirement, [175] but the temptation may remain.

Biologically plausible mechanisms for the development of disease can lend "credence to an inference of causality . . . [and depend] on the extent of scientific knowledge about the cellular and subcellular mechanisms through which the disease process works."[176] The error is for courts to *require* such knowledge. Requiring mechanistic data in addition to epidemiological data would commit Hill's error E5: *Requiring toxicology*.

For many adverse effects, the precise physiological cause is unknown and may be unknown for a considerable period of time. For instance, as of 1991 while the beneficial and harmful effects of aspirin were well known, scientists asserted, "We know the mechanism of action of some of these effects, but little about the mechanism of others."[177] A recent National Academy of Sciences committee noted that little is known about how statins cause muscle damage or thalidomide causes birth defects; it explained, "The history of science is replete with solid causal conclusions in advance of solid mechanistic understanding."[178] Sometimes researchers understand a few main mechanistic steps in a biological process leading to disease, but typically are unable to provide a detailed step-by-step identification of mechanistic events along the way from exposure to disease. At other times few mechanistic steps may be understood, yet there can be identification of the causes for a disease or dysfunction.

---

173. McClain v. Metabolife Int'l, Inc., 401 F.3d 1233 (11th Cir. 2005) (quoting Black v. Food Lion, Inc., 171 F.3d 308, 314 (5th Cir. 1999)); *see also* DeGidio v. Centocor Ortho Biotech, 3 F. Supp. 3d 674, 687 (citing *McClain*, 401 F.3d at 1253).

174. Jahn v. Equine Servs., PSC 233 F.3d 382, 390 (6th Cir. 2000) (attributing this view to the trial court).

175. Wagoner v. Exxon Mobil Corp., 813 F. Supp. 2d 771, 804 (E.D. La. 2011); *In re* Avandia Mktg., Sales Practices & Prods. Liab. Litig., No. 2007–MD–1871, 2011 WL 13576, at *1 (E.D. Pa. Jan. 4, 2011); Ferguson v. Riverside Sch. Dist. No. 416, No. CS–00–0097–FVS, 2002 WL 34355958, at *1 (E.D. Wash. Feb. 6, 2002).

176. Green et al., *supra* note 119, at 604.

177. TOXIC TORTS, 2d*, supra* note 77, at 179 (referencing Kenneth S. Santone & Garth Powis, *Mechanism of and Tests for Injuries*, *in* HANDBOOK OF PESTICIDE TOXICOLOGY 169 (W. J. Hayes Jr., & E. R. Laws Jr. eds., 1991).

178. NAT'L RESEARCH COUNCIL, REVIEW OF EPA'S INTEGRATED RISK INFORMATION SYSTEM (IRIS) PROCESS 85 (2014)

E. The Need for a Statistically Significant Association Before Applying Hill's Factors

A number of courts require that a statistically significant association between exposure and disease be established before Hill's factors may be used to assist causal inference.[179] Still other courts require that "an epidemiological study must be statistically significant at the 95% confidence level and the confidence interval may not include 1.0."[180] Other courts disagree with this constraint.[181]

The Reference Guide on Epidemiology notes, "In a number of cases, experts attempted to use these guidelines to support the existence of causation in the absence of any epidemiologic studies finding an association. . . . There may be some logic to that effort, but it does not reflect accepted epidemiologic methodology."[182]

This may not be appropriate *epidemiological* methodology, but it manifests error E4: R*equiring epidemiological evidence*. We raise this issue because courts—by asserting it—seem to suggest that Hill's considerations can only uniquely apply when litigants have statistically significant epidemiological studies. Importantly, however, if Hill's examinations can indeed assist causal inference, there seems to be no logical reason to restrict them only to human studies or only to statistically significant human studies, as long as there is an appropriate "association" that inspires inquiry. In fact, scientists beyond epidemiology have used actual or extended Hill's viewpoints to explore whether their data reveal causal relationships. Scientists from other biological sciences that certainly do not have human statistical data still use analogues of Hill's factors to assist causal inferences.[183] One group of scientists has even recommended

179. *See, e.g.*, *In re* Lipitor (Atorvastatin Calcium) Mktg., Sales Practices & Prods. Liab. Litig., 174 F. Supp. 3d 911 (D.S.C. 2016); Bearden v. Honeywell Int'l, Inc., No. 3:09–CV–1035, 2015 WL 7574344, at *1 (Nov. 23, 2015); *In re* Zoloft (Sertraline Hydrochloride) Prods. Liab. Litig., 26 F. Supp. 3d 449, 455 (E.D. Pa. 2014) (expert may not rely on nonstatistically significant studies to which to apply the BH factors); Mathews v. Novartis Pharm. Corp., No. 3:12–CV–314, 2013 WL 5780415, (S.D. Ohio 2013); Frischhertz v. SmithKline Beecham Corp., No. CIV.A. 10-2125, 2012 WL 6697124, at *1 (E.D. La. Dec. 21, 2012); *Wagoner,* 13 F. Supp. 2d at 803; Soldo v. Sandoz Pharm. Corp., 244 F. Supp. 2d 434, 569 (W.D. Pa.2003); Dunn v. Sandoz Pharm. Corp., 275 F. Supp. 2d 672 (M.D.N.C. 2003); Green et al., *supra* note 119.

180. Cano v. Everest Minerals Corp., 362 F. Supp. 2d 814, 821 (W.D. Tex. 2005) (citing Havner v. Merrell-Dow Pharm., 953 S.W. 2d at 717, 723 (Tex. 1997)).

181. *In re* Tylenol (Acetaminophen) Mktg., Sales Practices, and Prods. Liab. Litig., 181 F. Supp. 3d 278 (E.D. Pa. 2016) (finding that there is nothing saying a statistically-significant association must be found before applying the methodology); *In re* Zoloft (Sertraline Hydrochloride) Prods. Liab. Litig., No. 12–MD–2342, 2015 WL 314149, at *1 (E.D. Pa. Jan. 23, 2015).

182. Green et al., *supra* note 119, at 599 n.141 *referenced in Frischhertz*, 2012 WL 6697124, at *3.

183. *See generally* Richard B Lowell et al., *A Weight-of-Evidence Approach for Northern River Risk Assessment: Integrating the Effects of Multiple Stressors*, 19 ENVTL. TOXICOLOGY & CHEM, 1182 (2000).

modifications of Hill's factors in reviewing mechanistic data in an effort to determine whether it points to causation.[184] Importantly, a recent National Academy of Sciences committee has recommended that Hill's guidelines could be appropriately extended beyond human statistical studies.[185] And, a toxicological colleague of one of us (Professor Carl Cranor) confirms the extension of Hill's viewpoints well beyond epidemiology.[186] Thus, if courts have toxicologists or mechanistic experts that seek to use some of Hill's considerations but do not have epidemiological evidence, they should not summarily dismiss the experts for lacking epidemiological data, but instead attend to the cogency of the scientific argument in question based on appropriate associations.

F. Few Courts Recognize the Need for Sufficient Statistical Power

Once one is committed to the dichotomous thinking of tests of significance, or requiring that confidence intervals not include a null effect, one must be wary of studies whose population is so small that it is guaranteed to fail these tests. Alternative approaches have been proposed to bypass this problem.[187]

However, it appears that few courts recognize the need for statistical power in epidemiological studies to overcome this problem. This may be "widespread" because almost no courts seem sensitive to the statistical power of studies or even mention it in their opinions. When one does it stands out as unusual. The Federal Circuit Court of Appeals for the District of Columbia in *Ambrosini v. Labarraque* recognized that statistical studies needed sufficient power to detect adverse effects and uphold admission of an expert, Dr. Strom:

> Conventionally, in order to be considered meaningful, negative studies, that is, those which allege the absence of a causal relationship, must have at least an 80 to 90 percent chance of detecting a causal link if such a link exists; otherwise, the studies cannot be considered conclusive. Based on sample sizes too small to be reliable, the negative studies at issue, Dr. Storm explained, lacked sufficient statistical power to be considered conclusive.[188]

The good doctor's testimony was upheld by the court.

The IARC articulates how difficult it is properly to infer no adverse effects from studies that do not show adverse effects.

---

184. *See* M. E. (Bette) Meek et al., *Mode of Action Human Relevance (Species Concordance) Framework: Evolution of the Bradford Hill Considerations and Comparative Analysis of Weight of Evidence*, 34 J. APPLIED TOXICOLOGY 595 (2014).

185. NAT'L ACADS. OF SCIS., ENG'G & MED., *supra* note 22, at 7, 120–23 ("The causal guidelines that were developed by Bradford Hill (1965) . . . have proved particularly useful for interpreting epidemiological findings in the context of experimental and mechanistic evidence. Those guidelines have been proposed by others for evaluating adverse-outcome pathways (OECD 2013).").

186. Personal Communication from David A. Eastmond, Professor & Toxicologist, Dep't of Envtl. Toxicology, Univ. of Cal., Riverside, to Carl F. Cranor, Professor of Philosophy & Faculty Member Envtl. Toxicology Graduate Program, Univ. of California, Riverside (Feb. 15, 2017) [hereinafter Eastmond Personal Communication]; *see also* NAT'L ACADS. OF SCIS., ENG'G, & MED., *supra* note 22.

187. *See, e.g.*, Sander Greenland, *Nonsignificance Plus High Power Does Not Imply Support for the Null Over the Alternative*, 22 ANNALS EPIDEMIOLOGY 364 (2012).

188. Ambrosini v. Labarraque, 101 F.3d 129, 136 (D.C. Cir. 1996).

> When several epidemiological studies show little or no indication of an association between an exposure and cancer, the judgment may be made that, in the aggregate, they show evidence of lack of carcinogenicity. Such a judgment requires first of all . . . the possibility that bias, confounding or misclassification of exposure or outcome could explain the observed results should be considered and excluded with reasonable certainty. In addition, all studies that are judged to be methodologically sound should be consistent with a relative risk of unity for any observed level of exposure and, when considered together, should provide a pooled estimate of relative risk, which is at or near unity and has a narrow confidence interval, due to sufficient population size. Moreover, no individual study or the pooled results of all the studies should show any consistent tendency for relative risk of cancer to increase with increasing level of exposure. It is important to note that evidence of lack of carcinogenicity obtained in this way from several epidemiological studies can apply only to the type(s) of cancer studied and to dose levels and intervals between first exposure and observation of disease that are the same as or less than those observed in all the studies. Experience with human cancer indicates that, in some cases, the period from first exposure to the development of clinical cancer is seldom less than 20 years; latent periods substantially shorter than 30 years cannot provide evidence for lack of carcinogenicity.[189]

This constitutes a cautionary note about how to better assess when studies show "no adverse effect." Researchers and judges must exercise great care in judging that indeed there are no adverse effects.

Judges will need to exercise considerable care to ensure that, because of the epistemic norms implicit in science, their reviews protect as well against dismissing a true risk in error as against erroneously accepting as risky something that was safe.[190] Some courts seem to take "no effect" studies at face value. In *Chambers v. Exxon* the judge seemingly failed to note this point, excluding expert testimony simply because there were no statistically significant epidemiological studies showing benzene caused chronic myelogenous leukemia (CML).[191] Yet, at that time, cancer researchers were certain that benzene could cause CML.[192] The judge appeared to conclude that because there were no statistically significant epidemiological studies, a requirement he thought important, benzene did not cause CML. Subsequently, IARC has found that "although limited by low statistical power, the current meta-analysis provides support for a possible association of occupational exposure to benzene and the

189. INT'L AGENCY FOR RESEARCH ON CANCER, MONOGRAPHS ON THE EVALUATION OF CARCINOGENIC RISKS TO HUMANS § 2 (2006).

190. TOXIC TORTS, 2d*, supra* note 77, at 211.

191. Chambers v. Exxon Corp., 81 F. Supp. 2d 661, 665 (M.D. La. 2000) (citing *Brock v. Merrell-Dow Pharm., Inc.*, 874 F.2d 307, 312 (5th Cir. 1993) and noting that Exxon produced "a number of scientifically performed studies which demonstrate no association between exposure to benzene and development of CML [chronic mylogenous leukemia]," a quite rare disease that studies are unlikely to detect).

192. Eastmond Personal Communication, *supra* note 186.

risk of CML. Identifying studies of CML with elevated risks of 1.23, (CI): 0.93–1.63), 1.67 (95% CI: 1.02–2.74), 1.40 (95% CI: 0.86–2.27), and 1.68 (95% CI: 0.74–3.84 [respectively]."[193]

XI. SOME COURTS HAVE ACTUALLY READ HILL'S ADDRESS AND AVOIDED SOME OF THE ABOVE PITFALLS

A. Hill's Factors Are Neither Exhaustive nor Required

The *Milward* court observed, "Although [Bradford] Hill identified nine viewpoints, it is generally agreed that this list is not exhaustive and that no one type of evidence must be present before causality may be inferred."[194] Temporality is a necessary condition. Mere failure to satisfy some or all of "the BH criteria (sic) does not provide independent grounds for excluding expert testimony as unrealizable."[195]

B. Hill's Factors Are Not Necessary Conditions for Inferring Causation

Some courts correctly understand that Hill's factors are not *necessary conditions* to be satisfied for causal inferences.[196] Moreover, *The Reference Guide on Epidemiology* correctly notes:

> 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 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. Drawing causal inferences after finding an association and considering these factors requires judgment.[197]

C. Failure to Use or to Satisfy Hill's Factors Need Not Doom the Admission of an Expert

Although Hill's considerations can be quite helpful in inferring causation from epidemiological studies, some courts recognize that "failure to [use or to] satisfy the Bradford Hill criteria does not doom admission [of an expert] under

193. Jelle Vlaanderen et al., *Occupational Benzene Exposure and the Risk of Chronic Myeloid Leukemia: A Meta-Analysis of Cohort Studies Incorporating Study Quality Dimensions*, 55 AM. J. IND. MED. 779 (2012).

194. Milward v. Acuity Specialty Prods. Grp., 639 F.3d 11, 17 (1st Cir. 2011).

195. *In re* Viagra Prods. Liab. Litig., 572 F. Supp. 2d 1071 (D. Minn. 2008).

196. *See, e.g.*, Konrick v. Exxon Mobil Corp., No. CV 14-524, 2016 WL 439361, at *1 (E.D. La. Feb. 4, 2016); Cunningham v. Masterwear, Inc., No. 1:04-cv-1616-JDT-WTL, 2007 WL 1164832 (S.D. Ind. Apr. 19, 2007), Cook v. Rockwell Int'l Corp, 580 F. Supp. 2d 1071 (D. Colo. 2006); Blanchard v. Eli Lilly & Co., 207 F. Supp. 2d 308 (D. Vt. 2002); Magistrini v. One Hour Martinizing Dry Cleaning, 180 F. Supp. 2d 584 (D. N.J. 2002).

197. Green et al., *supra* note 119, at 600 *referenced in* Burst v. Shell Oil Co., No. CIV.A. 14-109, 2015 WL 3755953, at *1 (E.D. La. June 16, 2015).

Daubert."[198] Thus, although the Hill considerations have long been identified as one methodology for assisting the inference of causation in the law, it need not be the only one.

XII. SUMMARY AND CONCLUSIONS

Myriad lines of evidence can be scientifically relevant to judgments about toxicity, not simply statistically significant human data. There is no obvious hierarchy of evidence among them, and courts should not impose hierarchies. Scientists use inferences and good causal explanations to structure their reasoning, integrate different lines of evidence (by assessing the quality of the relevant and available data), and determine whether and how the different lines of evidence contribute to the conclusion. They should be permitted to do so for studies in support of legal testimony. It is up to judges to determine the reliability of testimony and its relevance to the issues of a case, and up to juries to determine how well scientists apply such reasoning in a particular case.

The observed values that result from Hill's recommended examinations can substantially assist causal inferences, but courts should refrain from treating his examinations like Koch's postulates. Instead, when faced with experts that use Hill's examinations, courts should appraise how the examinations have been applied and whether they support a sufficient inference, or a good causal story, for reliable testimony that is relevant to the case as required by *Daubert*. Therefore, in order to facilitate this effort we have presented the probabilistic foundation of Hill's examinations and discussed some errors that courts or commentators have exhibited.

198. *In re* Celexa and Lexapro Prods. Liab. Litig., 927 F. Supp. 2d 758 (E.D. Mo. 2013) (citing *In re* Viagra Prods. Liab. Litig., 658 F. Supp. 2d 950 (D. Minn. 2009); *In re* Neurontin Mktg., Sales Practices, & Prods. Liab. Litig., 612 F. Supp. 2d 116, 133 (D. Mass. 2009) (citing *In re Viagra*, 572 F. Supp. 2d at 1081).