# DIFFERENTIAL ETIOLOGY: INFERRING SPECIFIC CAUSATION IN THE LAW FROM GROUP DATA IN SCIENCE

Joseph Sanders,* David L. Faigman,** Peter B. Imrey,*** and Philip Dawid****

*In every toxic-tort case, the plaintiff must prove that the defendant exposed the plaintiff to something that caused an injury. The causal proof is in two parts: proof of general causation and proof of specific causation. General causation addresses the question of whether the exposure can cause injury in anyone. In the area of toxic torts, the evidence available to answer this question comes in the form of group-based studies in the fields of toxicology, epidemiology, and genetics—studies that search for the effects of causes.*

*Proof of general causation, however, is not enough. Because court cases focus on the individual, the plaintiff must prove specific causation. The plaintiff must show by a preponderance of the evidence that the exposure—and not something else—caused the plaintiff's harm. This typically is done through expert "differential etiology" testimony. This testimony is not focused on the search for the effects of causes but rather for the causes of effects.*

*Unfortunately, there is no body of science to which experts can turn when addressing this issue. Ultimately, much of the evidence that can be brought to bear on this causal question is the same group-level data employed to prove general causation. Consequently, the expert testimony often feels jerry-rigged, an improvisation designed to get through a tough patch. Court opinions that rule on the admissibility of this testimony rarely offer any systematic guidance about what is and is not an adequate differential etiology.*

---

* A. A. White Professor of Law, University of Houston Law Center.

** Chancellor, Dean, and John F. Digardi Distinguished Professor of Law, University of California Hastings College of the Law.

*** Professor of Medicine, Cleveland Clinic Lerner College of Medicine of Case Western Reserve University and Full Staff, Dept. of Quantitative Health Sciences, Cleveland Clinic Lerner Research Institute.

**** Emeritus Professor of Statistics, Cambridge University.

We wish to thank Michael D. Green, Steve Gold, and a workshop at the University of Houston for helpful comments on earlier drafts.

*Too often lawyers and courts have treated differential etiology as a matter of logical deduction and not as what it actually is: an inferential process combining statistical reasoning with a conceptual model of the causal interrelationships underlying observed data. This Article seeks to bring clarity to the differential-etiology determination by offering a scientifically oriented exposition of differential etiology. It provides a taxonomy of specific-causation situations and a systematic discussion of the factors that courts should consider in making a specific-causation determination. Our goal is to assist lawyers and judges in reasoning from group data to individual cases, what is referred to in the literature as the G2i problem.*

## TABLE OF CONTENTS

## INTRODUCTION

Off and on for several years, Ms. Jones took the drug Lipitor to control her blood-cholesterol level.[1] Apparently the drug was effective in this regard. During periods when she did not take the medication, her cholesterol levels rose and then fell again when she resumed taking the drug. However, several years after she began taking the drug, she was admitted to the hospital with a high blood-glucose level and diagnosed with new-onset diabetes.

Ms. Jones sued the manufacturer of Lipitor, claiming the drug caused her Type 2 diabetes. Plaintiff's proposed expert, a highly qualified individual with both an M.D. and a Ph.D. from prestigious universities, was prepared to testify both that Lipitor can cause diabetes and that Jones's diabetes was caused by Lipitor. The first opinion concerns the question of general causation. Can Lipitor cause diabetes? The second opinion concerns the question of specific causation. Did Lipitor cause Ms. Jones's diabetes?[2]

To answer the first question, Ms. Jones's expert could draw upon a potentially wide body of evidence. Scientists typically study groups of individuals with the general goal being to identify how these groups differ in relation to one or more variables. Hence, laboratory-based bioscientists might examine whether Lipitor and other statins cause diabetes in small animals bred for diabetes susceptibility, and epidemiologists might compare the rates of diabetes in human populations exposed to statins with others not similarly exposed.

But proof of general causation is not enough to resolve the legal issues. In the courtroom the focus is ultimately on the individual. The plaintiff must also prove by a preponderance of the evidence that the plaintiff's injury was caused by the substance in question. The plaintiff must find an expert prepared to conduct a

---

1. The facts in this introductory example are loosely taken from *In re Lipitor (Atorvastatin Calcium) Marketing, Sales Practices and Products Liability Litigation*, 150 F. Supp. 3d 644 (D.S.C. 2015) (relating to *Hempstead v. Pfizer, Inc.*, 2:14-cv-1879). We discuss this case at greater length below.

2. The distinction between "general causation" and "specific causation" is endemic to the science and law intersection and involves what has been termed the "G2i issue," or the challenge of reasoning from group data in science to what is typically most relevant in courts: whether an individual case is an instance of some general phenomenon. The challenges of G2i are explored in David L. Faigman et al., *Group to Individual (G2i) Inference in Scientific Expert Testimony*, 81 U. CHI. L. REV. 417 (2014).

"differential etiology"[3]—that is, to testify that it is more likely than not her injury was caused by the substance in question rather than the other possible causes of her injury.[4] In our example, Ms. Jones had a number of other "risk factors" for diabetes. She was of an age when people experience an increased risk of diabetes, she had a high body-mass index, one of her parents was diabetic, and she suffered from hypertension. What then can Ms. Jones's expert say about whether Lipitor caused her diabetes? What evidence can the expert bring to bear on this question?

The problem confronting the expert is well known. As we have noted in several previous articles, when courts turn from ascertaining general causation to specific causation, they are asked to ascertain not the effects of causes but the causes of effects.[5] However, in many tort cases and other cases of indeterminate causation, the additional science available to assist in this decision is quite limited, if it exists at all.[6] It is important to understand why this is the case. Practicing scientists tend to

---

3. Etiology is the study of causes. In medical usage, it is the study of the cause or origin of a disease. As discussed *infra* notes 26–28 and accompanying text, many courts use the term "differential diagnosis" when they are describing what is better understood as "differential etiology." For the sake of clarity, we use "differential etiology" to describe a method by which the cause of a condition is determined, in contrast to "differential diagnosis," which refers most accurately to the method by which the identity of the condition itself is determined. Hence, for example, the method for determining that the plaintiff suffers from new-onset diabetes is differential *diagnosis*; the method for determining that Lipitor caused that illness is differential *etiology*.

4. Most courts have resisted calls for alternatives such as permitting proportionate recoveries even when the plaintiff fails to meet the preponderance standard. Under such schemes, if a plaintiff could prove, for example, that her risk increased 25% because of her exposure to substance X, she could recover 25% of her damages from the maker of X. For discussions of various ways in which courts might alter causal proofs in mass exposure cases, see Alessandro Romano, *God's Dice: The Law in a Probabilistic World*, 41 U. DAYTON L. REV. 57, 84–86 (2016); David Rosenberg, *Individual Justice and Collectivizing Risk-Based Claims in Mass Exposure Cases*, 71 N.Y.U. L. REV. 210 (1996); David Rosenberg, *The Causal Connection in Mass Exposure Cases: A "Public Law" Vision of the Tort System*, 97 HARV. L. REV. 849, 898–900 (1984); Margaret Berger, *Eliminating General Causation: Notes Toward a New Theory of Justice and Toxic Torts*, 97 COLUM. L. REV. 2117 (1997); Margaret Berger & Aaron Twerski, *Uncertainty and Informed Choice: Unmasking* Daubert, 104 MICH. L. REV. 257, 275–80 (2005); Steve C. Gold, Note, *Causation in Toxic Torts: Burdens of Proof, Standards of Persuasion, and Statistical Evidence*, 96 YALE L.J. 376, 393–401 (1986). In this Article, our focus is on proof of specific causation, assuming the current liability regime. We do note that in most of the situations discussed in this Article, causation must be approached from a probabilistic perspective. *See* Frederick Schauer & Barbara Spellman, *Probabilistic Causation in the Law*, 176 J. INSTITUTIONAL & THEORETICAL ECON. 4 (2020).

5. *See* Faigman et al., *supra* note 2, at 435; *see also* David L. Faigman, Christopher Slobogin & John Monahan, *Gatekeeping Science: Using the Structure of Scientific Research to Distinguish Between Admissibility and Weight in Expert Testimony*, 110 NW. U. L. REV. 859, 888 (2016).

6. Most of the examples in this Article are taken from the area of what is commonly called "toxic torts." In this Article we use "toxic tort" to describe a wide variety of lawsuits. Included in our use of the term are not only the obvious examples of asbestos and benzene but also drugs that cause harm to their users. However, we also discuss cases that

address causal questions by elucidating and modulating the mechanisms that regularly produce an effect of interest, studying groups in order to control random variability. They do this much more often than by directly attempting to identify which, among proposed candidates, was responsible for a particular instance of that effect.[7] Thus, typically, the published scientific evidence to address such a question of specific causation is the same evidence a court must use to resolve general-causation questions.

Consequently, the expert testimony often feels jerry-rigged, an improvisation designed to get through a tough patch.[8] Court opinions that rule on the admissibility[9] of this testimony rarely offer any systematic guidance about what is and what is not an adequate differential etiology. Unfortunately, too often courts have viewed this inferential process as a matter of logical deduction and not as what it actually is: an inferential process combining statistical reasoning with a conceptual model of the causal interrelationships underlying observed data. This Article seeks to bring clarity to the differential-etiology determination. It provides scientifically informed guidelines for conducting an adequate differential etiology. As such,

---

would not generally be classified as toxic torts. For the purposes of this Article, the most important factor for inclusion is the possibility of multiple causes of the plaintiff's condition. The inferential reasoning challenge presented in toxic-tort cases is present in all areas of scientific evidence that involve reasoning from group data to individual cases, which is usually the ultimate issue of concern in court. Guns leave identifying marks on bullets, but the ultimate legal issue in court is whether the marks on the bullet that killed the victim came from the defendant's gun. Michael J. Saks & Jonathan J. Koehler, *The Individualization Fallacy in Forensic Science Evidence*, 61 VAND. L. REV. 199, 212–13 (2008). Hence, the challenge of reasoning from general data to the individual case is presented in virtually all areas in which science is applied in the courtroom. Faigman et al., *supra* note 2. Toxic torts are arguably the area in which courts first discovered the disconnect between the general data of science and the need for individuation in the law.

7. Epidemiological investigations of disease outbreaks and of the origins of new diseases and accident investigations of virtually any type are exceptions to this "rule" that fall in a gray zone between science and application.

8. One solution to this problem would be to prohibit all specific-causation testimony and simply leave the issue to the jury. *See* A. Philip Dawid et al., *Fitting Science into Legal Contexts: Assessing Effects of Causes or Causes of Effects?*, 43(3) SOCIO. METHODS & RSCH. 359, 363–64 (2014).

9. Our primary focus in this Article is on the federal courts. The current Federal Rule of Evidence 702 states:

> 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.

FED. R. EVID. 702.

however, these guidelines only direct a court's discretion; they do not prescribe outcomes.[10]

Our primary audiences for this Article are the lawyers and judges who must manage specific-causation testimony. Following *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, federal courts—and now most state courts—are charged with evaluating the methods and principles underlying proffered scientific evidence.[11] This holding was codified into amended Federal Rule of Evidence 702 in 2000. Whereas general causation is the subject of volumes of scientific discourse,[12] specific causation is a term largely unknown among scientists. Indeed, the method of differential etiology that is currently employed to assess specific causation was invented by lawyers and largely, if inadequately, defined by courts and legal scholars, not scientists.[13] This Article, then, is the first full-scale effort to bring scientific sensibilities—and rigorous statistical thinking—to the legally imperative concept of specific causation.

In Part I we begin with the historical background. We review the confusion that has surrounded the nature of differential-etiology testimony and the expertise of those offering it, and show how this confusion has impeded efforts to develop a systematic approach to understanding and analyzing this testimony. We end this Part by considering the steps one should take when deciding a toxic-tort case: (a) making a proper diagnosis; (b) supporting ("ruling in") the plausibility of the alleged cause of the injury on the basis of general evidence and logic; and (c) particularization, i.e., excluding ("ruling out") competing causes in the specific instance under consideration.[14]

Part II discusses problems that arise when trying to obtain a correct diagnosis of the illness for which a cause is sought. A correct diagnosis is a critical first step. If a diagnosis is incorrect, all other steps are irrelevant.

Part III discusses the evidence one may use to rule in the alleged cause of the plaintiff's injury. This is the general-causation question. What are the effects of an alleged cause? Although our focus is ultimately on the specific-causation question, i.e., the causes of a specific effect, this issue cannot be properly addressed in the absence of an understanding of what evidence is available on the general-causation issue and the internal- and external-validity problems confronting this evidence. The last portion of Part III uses the well-known Bradford Hill criteria to

---

10. The primary focus of this Article is to provide a road map to specific-causation decisions. But, in doing so we also address the current state of judicial evaluation of differential-etiology testimony.

11. *See* Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579 (1993); *see also* DAVID L. FAIGMAN ET AL., MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY § 1 (West 2019–2020 ed.).

12. Dawid et al., *supra* note 8, at 360.

13. David L. Faigman & Claire Lesikar, *Organized Common Sense: Some Lessons from Judge Jack Weinstein's Uncommonly Sensible Approach to Scientific Evidence*, 64 DEPAUL L. REV. 421, 439–40 (2015).

14. Best v. Lowe's Home Ctrs., Inc., 563 F.3d 171, 180 (6th Cir. 2009).

clarify how the different types of evidence available in these cases address or fail to address internal- and external-validity threats.[15]

Part IV turns to specific causation. Building on the discussion in Part III, it organizes specific-causation cases based on the evidence available to include the putative cause and to exclude other competing causes of a plaintiff's condition. The empirical task is always to assess the strength of the evidence supporting the asserted cause against the strength of the evidence supporting all competing causes. Our goal is to provide a taxonomy of specific-causation situations and provide a systematic discussion of the factors that courts should consider in making a specific-causation determination. As we shall see, the evidence for specific causation is often the same evidence used to assess general causation. Therefore, although the Bradford Hill guidelines were originally put forth as indicia by which to establish general causation, together with close evaluation of validity, they may also assist the effort to evaluate specific causation.

Part V considers the complementary roles of science and policy in guiding admissibility decisions. The approach set forth in Part V can help courts to understand and articulate the grounds on which they admit or exclude expert evidence. It is important to state, however, that they cannot establish a fixed standard for admissibility. The standard is, or at least should be, grounded in science; however, it is also influenced by policy considerations. These considerations influence how much scientific support a specific-causation opinion should have before it is admitted into testimony. In Part V we acknowledge this fact. We do not attempt to take sides on how much evidence is needed, but we do urge courts to do their best to always explain their decisions using the approach developed in Part V. This Article closes with a brief conclusion.

## I. HISTORICAL BACKGROUND

The terminology used to describe the process of establishing specific causation has been clouded by ambiguity. Courts frequently describe this as a process of "differential diagnosis." Differential diagnosis has a well-understood meaning in medicine. Stedman's Medical Dictionary defines the term as "the determination of which of two or more diseases with similar symptoms is the one from which the patient is suffering by a systematic comparison and contrasting of the clinical findings."[16] In other words, it is a process by which physicians determine the patient's condition, not the cause of that condition.

It is important to understand that the causal analysis associated with a diagnosis has itself changed over time and, indeed, may change with respect to a given disease as our understanding of that disease improves. Clinical diagnosis originated as a simple recognition and labeling of patterns of symptoms and signs. Causal concepts such as humours, miasmas, and spirits were vague, pre-scientific, and of little utility for prevention. However, since the discovery of microbes and revolutions in biochemistry and genetics, mechanistic concepts have worked their

15. Austin Bradford Hill, *The Environment and Disease: Association or Causation?*, 58 PROC. ROYAL SOC'Y MEDICINE 295 (1965).

16. *Differential Diagnosis*, STEDMAN'S MEDICAL DICTIONARY (27th ed. 2000).

way into disease definitions.[17] Nowadays, diagnosis of an infectious disease typically relies not only on characteristic symptoms and signs of illness but also on identifying a microorganism, a uniquely associated antigen, or minimally, antibodies that establish that the organism is or was present.[18] Identifying a cancer depends on observations at the histological, cellular, and sometimes genetic levels. But these mechanisms that now determine the diagnosis are actually preceding causes of the signs and symptoms of which diagnoses used to be entirely constituted. So it is somewhat natural that, as the concept of diagnosis has moved from pure cross-sectional observation without any time dimension to identification of preceding active mechanisms these signs and symptoms ultimately manifest, the terminology of purely cross-sectional labeling has persisted.

Consider how the definition of AIDS has evolved, from labeling of a bewildering, unexplained clinical syndrome or syndromes initially, to a disease defined by low prevalence of CD4 cells or onset of one of a set of more specific diseases preceded or accompanied by identification of human immunodeficiency virus (HIV) using antigen, antibody, or RNA tests.[19] So, HIV infection has been clearly identified as the cause of AIDS and is now inextricably part of the meaning of that diagnosis. Diagnosis has become more etiologic.

Consider another example where the two tasks overlap. In the process of diagnosing a patient's ailment, allergists routinely search for the cause of an individual's allergy by exposing the individual to various substances and observing the results. In this and similar situations, the difference between diagnosis and etiology shrinks.

Even in this evolving environment, however, one important difference remains between the physician's task and the law's objective. In most clinical situations, the physician wants to understand the causal mechanisms producing an ailment only to the point of being able to prescribe a course of treatment (or to determine that there is no available treatment). Usually, a full understanding of *background causes* is not necessary to achieve this objective.[20] But it is these very background causes that are of interest to the legal system when it conducts a differential etiology. To return to the AIDS example, in a legal context when one

---

17. *See generally* ROY PORTER, THE GREATEST BENEFIT TO MANKIND: A MEDICAL HISTORY OF HUMANITY (1997); STEVEN JOHNSON, THE GHOST MAP: THE STORY OF LONDON'S MOST TERRIFYING EPIDEMIC—AND HOW IT CHANGED SCIENCE, CITIES, AND THE MODERN WORLD (2006).

18. *See* CONTROL OF COMMUNICABLE DISEASES MANUAL (David L. Heymann ed., 20th ed. 2014).

19. ROGER C. HERDMAN & CLYDE J. BEHNEY, OFF. OF TECH. ASSESSMENT, THE CDC'S CASE DEFINITION OF AIDS: IMPLICATIONS OF THE PROPOSED REVISIONS II 53–II 56 (1992), https://ota.fas.org/reports/9206.pdf [https://perma.cc/JB8U-UJQY]; OFF. OF AIDS RSCH., DEP'T OF HEALTH & HUM. SERVS., GLOSSARY OF HIV/AIDS-RELATED TERMS 2 (9th ed. 2021) https://clinicalinfo.hiv.gov/sites/default/files/glossary/Glossary-English_HIVinfo.pdf [https://perma.cc/P4G2-5SPZ].

20. We have chosen to use the term "background causes" to describe the causes of interest to the law. Often, but not always, they are distal rather than proximal causes, but they share in common that many are causes that are relevant to the legal task of attributing responsibility.

asks about the cause of a patient's disease, it may often be taken for granted that what is meant is not HIV itself but the cause of the patient's HIV infection, be that one type or another of sexual contact, illicit drug injection, contaminated blood transfusion, or other inadvertent exposure of someone during the health-care process.

There are important exceptions to this general position. For example, as we discuss below, a physician's search for a genetic cause of an ailment may be a search for the same background causal explanation that is of interest to the law. In most cases, however, the causal analysis of interest to the law is not one that physicians are trained to perform.

Beginning at least as early as the 1940s, courts used the term "differential diagnosis" to describe the diagnostic process.[21] It is difficult to determine when the term "differential diagnosis" first appeared in legal opinions to describe the background cause that explains the plaintiff's disease. Candidates can be found as early as the mid-1960s,[22] but clear use of the term in this way does not arise prior to the 1980s.

The earliest cases involved vaccines.[23] For example, in *Grill v. United States*, Judge Pratt noted that the plaintiff's expert had relied on a differential diagnosis which "ruled out other possible causes" in support of his conclusion that the likelihood that the plaintiff's "optic neuritis was caused by the swine flu vaccine exceeds ninety percent."[24]

This usage escaped the bounds of the vaccine cases with the rise of toxic torts and the increasingly frequent challenges to the admissibility of the plaintiff's expert opinions in the early 1990s.[25] The number of cases steadily grew following the U.S. Supreme Court's *Daubert* opinion.

Early cases and, to a lesser extent, cases still today exhibit three shortcomings in their approach. First, the courts failed to distinguish between differential diagnosis as it is understood in the medical field and the search for the legally relevant background cause of an illness. Second, because of this confusion, the courts failed to recognize that expertise in arriving at a correct diagnosis does not necessarily transfer to expertise in assigning causes to illnesses. And third, perhaps in part because of these two confusions, courts failed to develop a coherent, explicit "standard technique" one should follow when one attempts to answer the background causal question.

---

21. Phoenix Mut. Life Ins. Co. of Hartford, Conn. v. Harmegnies, 110 F.2d 20, 26 (8th Cir. 1940). Most cases that use the term in this way are medical-malpractice claims based on the allegation that the physician failed to properly diagnose the plaintiff's illness.

22. *See* Martin v. Westchester Fire Ins. Co., 183 So. 2d 769 (La. Ct. App. 1966).

23. *E.g.*, *In re* Swine Flu Immunization Prod. Liab. Litig., 533 F.Supp. 703, 710 (D. Utah 1982).

24. 552 F. Supp. 505, 507 (E.D.N.Y. 1982). The court ultimately excluded this testimony, finding that it was based on little more than temporal proximity between the vaccine and the onset of symptoms.

25. *See* Hines v. Consol. Rail Corp., 926 F.2d 262, 273 (3d Cir. 1991); *In re* Paoli R.R. Yard PCB Litig., 35 F.3d 717, 755 (3d Cir. 1994).

In the ensuing years, many, but not all, courts have exhibited a greater appreciation of the first and second mistake, but courts have made relatively little progress in developing a systematic analytical approach. Before we turn to such an approach, however, it is worthwhile to say a bit more about how the first two misunderstandings have impeded the courts' ability to deal with the third failure.

### *A. Recognizing that Differential Etiology and Differential Diagnosis are Different Endeavors*

In the 1995 case of *McCullock v. H.B. Fuller Co.*,[26] the court used the term "differential etiology" for what appears to be its first recorded use. Although the *McCullock* court did not appear to appreciate the significance of this change in phraseology, differential etiology is a better choice than differential diagnosis because it contemplates the study of causes or origins. Although many courts now distinguish between differential diagnosis and differential etiology,[27] some continue to confuse the two terms.[28]

At one level, using the term "differential diagnosis" to identify specific causation is only semantic. However, at another level the confusion can mislead courts into believing that physicians are always well trained in this process. In this Article we refer to the process of determining specific causation as "differential etiology."

### *B. Recognizing that Expertise in Assigning Illnesses to Symptoms Does Not Transfer to Expertise in Assigning Causes to Illnesses*

The use of the term "differential diagnosis" to mean the search for the background cause of an illness has led some courts to begin their analysis of a specific-causation expert with an assertion that this "method" is widely accepted and practiced by physicians.[29] An aura of validity and general acceptance over anything going by that name causes courts to be less critical than they might be if they recognized that skill and experience in diagnosing an illness does not always translate to skill and experience in attributing background causation.[30] For most physicians, attributing background causation is not part of their normal practice.[31]

---

26. 61 F.3d 1038, 1043 (2d Cir. 1995).

27. *See, e.g.*, McClain v. Metabolife Int'l, Inc., 401 F.3d 1233, 1252 (11th Cir. 2005); Hendrix *ex rel.* G.P. v. Evenflo Co., 609 F.3d. 1183, 1195 & n.5 (11th Cir. 2010). *See generally* Anthony G. Hopp et al., *Differential Diagnosis and* Daubert*: Preventing the Misuse of Differential Etiology to Prove Causation in Toxic Tort Cases*, 84 DEF. COUNS. J. 1 (2017).

28. See, for example, the definition offered by the court in *Glastetter v. Novartis Pharmaceuticals Corp.*, 107 F. Supp. 2d 1015, 1019 (E.D. Mo. 2000). "Differential diagnosis is a patient-specific process of elimination that medical practitioners use to identify the 'most likely' cause of a set of signs and symptoms from a list of possible causes." Lennon v. Norfolk & W. Ry. Co., 123 F. Supp. 2d 1143 (N.D. Ind. 2000).

29. *See* Benedi v. McNeil-P.P.C., Inc., 66 F.3d 1378, 1384 (4th Cir. 1995); John's Heating Serv. v. Lamb, 46 P.3d 1024, 1036 (Alaska 2002); Hyman & Armstrong, P.S.C. v. Gunderson, 279 S.W.3d 93, 107 (Ky. 2008).

30. *See* Hines v. Consol. Rail Corp., 926 F.2d 262, 273 (3d Cir. 1991); Ervin v. Johnson & Johnson, Inc., 2006 WL 1529582, at *4–5 (S.D. Ind. May 30, 2006).

31. Erica Beecher-Monas, *Lost in Translation: Statistical Inference in Court*, 46 ARIZ. ST. L.J. 1057, 1058 (2014); *see also* Dawid et al., *supra* note 8, at 367.

Distinguishing etiology from diagnosis ensures that courts do not assume that all physicians are trained to engage in a differential etiology or that this is part of their routine practice.[32]

*Bowers v. Norfolk Southern Corp.* [33] illustrates another important difference. Judge Royal begins by distinguishing between traditional diagnoses and establishing the background cause of a disease, and he then focuses his attention on the consequences of a mistaken diagnosis versus a mistaken courtroom etiology. A missed or mistaken diagnosis may lead to a patient's death and, as a result in some cases, a malpractice suit.[34] Even when there is no question of malpractice, a misdiagnosis presumably leads to an ineffective treatment, causing the physician to reconsider the original diagnosis. That is, there is a natural self-correcting feedback loop that assists physicians in the process of abandoning incorrect diagnoses. Because of this feedback loop, physicians, through long experience in making diagnoses, may well become experts in that process.[35] There is no such feedback loop for differential-etiology testimony in court. Moreover, the expert witness has little at stake. The plaintiff is at no risk of harm beyond the loss of the case, and the expert will not be sued for malpractice.[36] Judge Royal concludes:

> The differential diagnosis method has an inherent reliability; the differential etiology method does not. This conclusion does not suggest that the differential etiology approach has no merit. It simply means that courts, when dealing with matters of reliability, should consider opinions based on the differential etiology method with more caution. It also means that courts should not conflate the two definitions.[37]

### *C. Developing a More Coherent and Explicit Procedure a Court Should Follow When It Attempts to Answer the Specific-Causation Question*

Clarity on the difference between routine diagnostic practices and differential etiology as the term is used by the law clears away the potential for confusion and bias that arises from conflating the two concepts. However, it does not tell us how to go about assessing an expert's search for background causes of an effect.

---

32. *See* Thomas v. Novartis Pharms. Corp., 443 Fed. App'x. 58, 62 (6th Cir. 2011).

33. 537 F. Supp. 2d 1343, 1359–60 (M.D. Ga. 2007).

34. *See, e.g.*, Rhodes v. United States, 967 F. Supp. 2d 246 (D.D.C. 2013) (physician failed to consider breast cancer as part of her differential diagnosis of patient).

35. For a useful discussion of expertise acquired in this way, see David L. Faigman et al., *Check Your Crystal Ball at the Courthouse Door, Please: Exploring the Past, Understanding the Present, and Worrying About the Future of Scientific Evidence*, 15 CARDOZO L. REV. 1799 (1994). For a discussion of the importance of the frequency and timing of feedback in assessing expert judgment and intuition, see Daniel Kahneman, *Chapter Twenty-Two*, *in* THINKING FAST AND SLOW (2011).

36. FAIGMAN ET AL., *supra* note 11, § 2.

37. Bowers, 537 F. Supp. 2d at 1361. *See generally* David L. Faigman & Jennifer Mnookin, *The Curious Case of* Wendell v. GlaxoSmithKline, LLC, 48 SETON HALL L. REV. 607, 614 (2018).

Courts have offered very little systematic guidance on this point. Many courts have reiterated the position first set out in *In re Paoli R.R. Yard PCB Litigation*.[38] For example, *Best v. Lowe's Home Centers, Inc.* relies on that case in offering up the following specific-causation guidance to lower courts in the Sixth Circuit:

> We hereby adopt the following differential-diagnosis test, adapted from the Third Circuit's well-reasoned opinion: A medical-causation opinion in the form of a doctor's differential diagnosis is reliable and admissible where the doctor (1) objectively ascertains, to the extent possible, the nature of the patient's injury ("A physician who evaluates a patient in preparation for litigation should seek more than a patient's self-report of symptoms or illness and . . . should . . . determine that a patient is ill and what illness the patient has contracted."), (2) "rules in" one or more causes of the injury using a valid methodology, and (3) engages in "standard diagnostic techniques by which doctors normally rule out alternative causes" to reach a conclusion as to which cause is most likely.[39]

The *Best* court calls this a "test," but, in fact, it is only an outline that describes the steps required if one is to conduct a differential etiology. Step One requires the expert to properly ascertain the injury, i.e., perform a differential diagnosis in the medical sense. Step Two requires the expert (or often another expert) to determine whether the exposure in question is capable of causing the claimed condition. Only if a relevant dose of the substance could cause the condition should the claimant be able to prove that it did so in a particular case. Thus, the second step is to "rule in" the putative cause. Step Three is the critical step of "ruling out" alternative causes.

The following three Parts of this Article follow this roadmap: getting the diagnosis right, ruling in suspect causes (general causation), and ruling out competing causes (specific causation). Although the central purpose of this Article is to propose detailed guidelines to assess specific causation through the method of differential etiology, this journey necessarily goes through general causation. First and foremost, if there is no proof of general causation, the analysis must end there; claims of specific causation, *ipso facto*, cannot proceed without proof of general causation. Second, the evaluation of specific causation depends necessarily on the quantum of proof of general causation, not only as to the putative cause of the plaintiff's condition but also as to alternative causes for that condition. Proof of specific causation is invariably bound up with the available evidence for general causation for all potential causes.

## II. GETTING THE DIAGNOSIS RIGHT

The first step of differential etiology is coming to a proper diagnosis. Occasionally, what may appear to be a disagreement about the cause of the individual's injury is a disagreement about the diagnosis. This has occurred in the

38. *In re* Paoli R.R. Yard PCB Litig., 35 F.3d 717 (3d Cir. 1994).
39. 563 F.3d 171, 180 (6th Cir. 2009).

asbestos litigation because if an individual has an ailment strongly linked to asbestos exposure—asbestosis or mesothelioma—then the cause of the ailment is not in doubt. A diagnosis of mesothelioma is rarely controversial. However, the parties sometimes disagree as to whether the individual is suffering from asbestosis or from some other lung impairment.[40] Diagnostic disputes arise with respect to other exposures as well.[41] Courts rarely, if ever, exclude an expert's diagnosis testimony. However, they may well exclude his etiology testimony at least in part because of the shaky diagnosis.[42]

Although relatively few cases presently turn on the diagnosis, the increasing use of genetic information may cause this situation to change. *Bowen v. E.I. du Pont de Nemours and Co.* is one case where this was true.[43] In *Bowen*, the plaintiff and several other children were exposed to the fungicide Benlate in utero, allegedly causing birth defects. The defendant claimed, however, that the plaintiff suffered from CHARGE syndrome, a genetic ailment involving multiple birth defects. Plaintiff's experts disagreed because they believed the plaintiff did not meet enough of the criteria for this diagnosis. Therefore, they ruled out genetics as a cause. However, the court acquiesced to a defense request for genetic testing. The results indicated that the plaintiff's genetic profile contained a gene, CHD7, that had mutated. To that point in time, while not all individuals diagnosed with CHARGE syndrome exhibited this mutation, every individual with this mutation had been diagnosed as having the ailment. Ultimately, the trial court excluded the testimony of one of the plaintiff's experts, Dr. Howard, who, in spite of this new information, continued to argue that Benlate caused the plaintiff's birth defects and that in some unexplained way the fungicide and the mutation acted together to cause her injury.[44]

*Bowen* represents an extreme case because, according to the court, everyone with the mutated gene ultimately suffers from the syndrome. It is rare for any single genetic defect to inevitably cause a specific injury. Nevertheless, genetic information has become more important in narrowing the possible causes of an individual's ailment. We discuss the potentially increasing importance of genetic information below.

## III. RULING IN: GENERAL CAUSATION

Although the central goal of this Article is to help decision makers with the problem of specific causation, such arguments are contingent on a general-causation framework. As the following discussion of validity threats, types of evidence, and

---

40. *See, e.g.*, Borg-Warner Corp. v. Flores, 232 S.W.3d 765, 766–68 (Tex. 2007).

41. *Tamraz v. Lincoln Electric Co.*, 620 F.3d. 665 (6th Cir. 2010) involved a dispute as to whether the plaintiff was suffering from Parkinson's Disease or manganism. As in the asbestos situation, the diagnosis is critical to the issue of causation as the latter disease is caused by exposure to manganese, a substance with which the plaintiff came in contact as a welder.

42. This was the outcome in *Tamraz*, *id.* at 670.

43. Civ. A. 97C-06-194 CH, 2005 WL 1952859 (Del. Super. Ct. June 23, 2005), *aff'd*, 906 A.2d 787 (Del. 2006).

44. The Delaware Supreme Court affirmed the Superior Court's summary judgment in favor of the defendant on different grounds. Bowen v. E.I. du Pont de Nemours & Co., Inc., 906 A.2d 787 (Del. 2006).

the Bradford Hill indicia suggests, the evidence for general causation with respect to an exposure in question—exposure meaning a toxicological, microbiological, environmental, pharmaceutical, or other suspected disease-promoting agent—may be relatively weak or quite strong. Other things being equal, when the evidence for general causation is strong, and especially when the strength of the exposure–disease relationship as demonstrated in a body of research is substantial, the plaintiff faces a lower threshold in establishing the substance as the cause in a particular case than when the relationship is weaker. Only with a good understanding of the evidence available to prove general causation can one develop a systematic approach to the determination of specific causation. Moreover, in some situations the scientific evidence used to establish general causation is in fact the primary evidence available to establish specific causation.[45] We thus begin with a discussion of the second step of the analysis: establishing general causation, on which subsequent consideration of specific causation will often rest.

Evidence of causation, whether general or specific, is often a matter of judgment. Because they are frequently referred to in toxic-tort opinions, we use the well-known Bradford Hill indicia as one approach for assessing general causation.[46] Bradford Hill wrote for a sophisticated, scientific audience and was primarily concerned with determining general causation—what he would have simply referred to as causation. Our objective in this Part is to place the indicia in context. We endeavor to make certain assumptions explicit that Bradford Hill took for granted. Specifically, we consider the bread-and-butter scientific issues of validity (both internal and external). Moreover, although Bradford Hill was only addressing causal inference in the field of epidemiology, we use the indicia to discuss inferences in toxicology, clinical trials, and the emerging field of genetics. This Part explores the myriad factors that scientists would consider when coming to a judgment regarding general causation. Therefore, we defer an explication of the Bradford Hill indicia until after a discussion of validity and the types of evidence used in assessing general causation.

We also note here that in the 56 years since Bradford Hill's seminal publication, his indicia of causation have been repeatedly reappraised in the contexts of their applicability to specific problems and have been subject to a number of

---

45. See *infra* Section IV.

46. Some courts have approached Bradford Hill's article as setting forth *criteria* for general causation. This is an over-reading of his intentions. We agree with Susan Haack's comments concerning the limits of what Bradford Hill provided. She, and we, prefer the term *indicia* (or *guidelines* or perhaps *touchstones*) to *criteria*. She notes they are not criteria in the sense that they provide a decision procedure or a checklist to be mechanically followed. Susan Haack, *Proving Causation: The Holism of Warrant and the Atomism of* Daubert, 4 J. HEALTH & BIOMEDICAL L. 253, 276 (2008). Ultimately, a conclusion about causation in close cases is a matter that is difficult if not impossible to reduce to an algorithm. We will be satisfied if we provide a roadmap that allows courts to more clearly articulate the types of information upon which such judgments should be based.

The Bradford Hill indicia are far from the final word in assessing causation. In the last few decades, statisticians, epidemiologists, and others have developed a number of other sophisticated methods of assessing general causation. *See, e.g.*, JUDEA PEARL & DANA MACKENZIE, THE BOOK OF WHY: THE NEW SCIENCE OF CAUSE AND EFFECT (2018).

scientific and philosophical critiques.[47] During the intervening period the mathematical, statistical, philosophical, and computer-science communities have engaged with problems of general and specific causal inference, including the development of more formal, mathematical languages for modeling causality than existed during Bradford Hill's lifetime.[48] As the field of causal analysis has matured, these developments are increasingly appearing in monographs, textbooks, and more recently, popularizations for wider audiences, some of which suggest legal-causation relevance.[49] While this work has not and indeed may never converge into a single, generally accepted "science of causation," some of its concepts will likely be introduced into tort proceedings and ultimately influence tort law in ways we cannot now foresee. Rather than anticipate the specifics of that future, we will assume change will be evolutionary and here argue that the Bradford Hill indicia, whose utility for appraising general causation is generally accepted in tort law, can contribute to the resolution of specific-causation questions under current tort law in ways unlikely to be dramatically overthrown. We might add that while undoubtedly this literature will eventually influence how courts think about general causation, its application to the specific-causation problem is still in very early stages.

We begin our discussion of general causation with a topic largely implicit in the Bradford Hill Guidelines: threats to validity. We then review the types of data that may be brought to bear on the general causal question, including toxicological, epidemiological, and genetic data.

### *A. Validity*

A fundamental objective of science is to make causal assertions based on reproducible empirical relationships. However, empirical observations may mislead

---

47. *See* Nancy Potischman & Douglas L. Weed, *Causal Criteria in Nutritional Epidemiology*, 69(6) AM. J. CLINICAL NUTRITION 1309S (1999); Michael Höfler, *The Bradford Hill Considerations on Causality: A Counterfactual Perspective*, 2 EMERGING THEMES EPIDEMIOLOGY 11 (2005); Thomas A. Glass et al., *Causal Inference in Public Health*, 34 ANN. REV. PUB. HEALTH 61 (2005); Kristen M. Fedak et al., *Applying the Bradford Hill Criteria in the 21st Century: How Data Integration Has Changed Causal Inference in Molecular Epidemiology*, in 12 EMERGING THEMES IN EPIDEMIOLOGY 14 (2015); John P. A. Ioannidis, *Exposure-Wide Epidemiology: Revisiting Bradford Hill*, 35(11) STATS. IN MED. 1749 (2016).

48. *See* ANDREW GELMAN & JENNIFER HILL, DATA ANALYSIS USING REGRESSION AND MULTILEVEL/HIERARCHICAL MODELS (2007); JUDEA PEARL, CAUSALITY: MODELS, REASONING, AND INFERENCE (2d ed. 2009); CAUSALITY: STATISTICAL PERSPECTIVES AND APPLICATIONS (Carlo Berzuini, Philip Dawid & Luisa Bernardinell eds., 2012); TYLER VANDERWEELE, EXPLANATION IN CAUSAL INFERENCE: METHODS FOR MEDIATION AND INTERACTION (2015); JOSEPH Y. HALPERN, ACTUAL CAUSALITY (2019); PAUL ROSENBAUM, OBSERVATION AND EXPERIMENT: AN INTRODUCTION TO CAUSAL INFERENCE (2019); MARIUSZ MAZIARZ, THE PHILOSOPHY OF CAUSALITY IN ECONOMICS: CAUSAL INFERENCES AND POLICY PROPOSALS (2020); ALEXANDER BOCHMAN, A LOGICAL THEORY OF CAUSALITY (2021).

49. *See, e.g.*, PEARL & MACKENZIE, *supra* note 46.

us, and the causal conclusions we draw from the observations may be invalid due to a study's data collection plan and/or data-analytic approach.

Formally, scientific research typically examines relationships between putative causes and effects of interest in collections (i.e., "samples" of persons, animals, tissue, or cells with defined characteristics) under stipulated study conditions—these being the circumstances of the research in the laboratory, clinic, field, or other specialized research environment—and uses these observed relationships to draw inferences about broader "populations" or processes. Such inferences, whether causal or only correlational, are considered to be "internal" if they are generalizations of relationships seen in particular studies to larger populations with the same or very similar defined characteristics, under the same or very similar conditions of those studies, or alternatively, "external" if they are extrapolations to notably different populations and/or conditions. Validity of a research study refers to the extent to which a study's methods preclude or diminish potential validity threats—i.e., flaws in data collection and/or analysis that distort such inferences.[50] The validity of a study's conclusions is measured by the extent to which the study has successfully excluded plausible alternatives to its claimed exposure–disease relationships.

In the context of this Article, validity inquiries involve distinguishing (a) exposure—condition associations for which no such validity threats are sufficiently plausible to viably compete with the causal hypothesis, from (b) exposure–condition relationships that do admit viable alternative explanations because one or more validity threats cannot be confidently excluded.

Validity is a continuous assessment but may be described on a graded scale, e.g., low, high, very high, or invalid and valid when abbreviated at the extremes. "Internal validity" and "external validity" refer, respectively, to the validity of internal and external inferences. We discuss these in turn, giving greatest attention to the former because external validity is necessarily contingent on internal validity.

*1. Internal Validity*

Internal validity of a causal inference is demonstrated by exclusion of specific threats to that inference, i.e., noncausal or unreliable explanations for empirical associations based on a study's methods rather than reproducible relationships in nature. With few exceptions (e.g., biological placebo effects), validity threats fall within five broad categories: time reversals, selection bias, information bias, confounding, and chance.[51] We briefly introduce each.

---

50. DAVID G. KLEINBAUM ET AL., EPIDEMIOLOGIC RESEARCH: PRINCIPLES AND QUANTITATIVE METHODS 181–280 (1982) (discussing validity of epidemiologic research); WILLIAM R. SHADISH ET AL., EXPERIMENTAL AND QUASI-EXPERIMENTAL DESIGNS FOR GENERALIZED CAUSAL INFERENCE (2002).

51. Daniel I. Sessler & Peter B. Imrey, *Clinical Research Methodology 1: Study Designs and Methodologic Sources of Error*, 121 ANESTHESIA & ANALGESIA 1034, 1035–41 (2015); Daniel I. Sessler & Peter B. Imrey, *Clinical Research Methodology 2: Observational Clinical Research*, 121 ANESTHESIA & ANALGESIA 1043, 1046–47 (2015); Aalok K. Kacha et al., *Clinical Study Designs and Sources of Error in Medical Research*, 32 J. CARDIOTHORACIC & VASCULAR ANESTHESIA 2789, 2790–94 (2018); KLEINBAUM ET AL., *supra* note 50.

All definitions of cause-and-effect relationships require cause to precede effect in time. Time reversal occurs when research studies misconstrue the time sequence of exposure and condition. This seems unlikely in principle but in fact can easily happen when, as is usually the case, the specific timing of exposure onset and condition initiation are unknown. Consider, for instance, a hypothesized microbial cancer initiator. Researchers comparing tissue samples from cancer patients with tissue from comparable sites in healthy controls might base a causal inference on more frequent recovery of the microbe from tissue of cancer patients than from other patients with noncancerous conditions. However, an alternative explanation is that changes in the tissue environment due to preclinical cancer, such as a change in pH, may have made formerly healthy tissue more subject to colonization by the subject microbe. The cancer might thus be a contributory cause to the organism's ability to colonize the tissue; instead of the organism causing the cancer, the cancer may have caused colonization by the microbe. More generally, such time reversals are frequently at least conceivable when an exposure occurs gradually over time, and as for cancer generally, the time of onset of a pathological process manifesting later as disease is unknown. Such a biologically plausible explanation would greatly diminish the internal validity of the causal inference.[52]

Selection bias refers to a tendency of a study's methods to either understate or overstate the exposure–disease association in the target population due to an asymmetric subject-selection process that gives different chances of study inclusion to different individuals and members of different subgroups. Two well-known examples are Berkson's bias and regression to the mean. In an early example of the former,[53] the relationship between cigarette smoking and lung cancer was apparently attenuated in a case-control comparison by conducting the study within hospitals, where patient smoking exposures were generally elevated due to then-unknown relationships of smoking to cardiovascular and other diseases. Patients required extended hospitalizations at a time when acute myocardial infarctions from cardiovascular diseases were initially treated with weeks of hospitalization, so that smoking exposure in hospitalized patients tended to overrepresent smoking in the general population. Regression to the mean occurs when individuals are selected for some intervention or subsequent assessment based on their extreme values of what is to be assessed. Because data often fluctuate randomly due to inherent variability of target quantities or random-measurement error, we should expect measures of group centrality to become less extreme when measured subsequently due simply to the play of chance. For example, if we select students who did poorly on the LSAT

52. For a case that turned in part on this issue, see *Guinn v. AstraZeneca Pharmaceuticals LP*, 602 F.3d 1245, 1254 (11th Cir. 2010). The plaintiff averred that the drug Seroquel, prescribed to treat her bipolar depression, caused weight gain, which in turn caused her to develop diabetes. Although the plaintiff was not diagnosed with diabetes until four years after beginning to take Seroquel, the court noted that the development of diabetes occurs gradually over many years, and approximately half the patients diagnosed with diabetes have actually had the disease for about five years prior to diagnosis. Thus, "the temporal relationship in this case does not provide strong evidence of causation; in fact, it appears to equally indicate that Guinn may have already developed diabetes before ever taking Seroquel."

53. Richard Doll & Austin Bradford Hill, *Smoking and Carcinoma of the Lung: Preliminary Report*, 2 BRITISH MED. J. 739, 743–44 (1950).

for an LSAT-prep class, their subsequent improved aggregate performance could be due to the benefits of the class or simply regression to the mean of the individuals' abilities. Similarly, regression to the mean explains why last year's best mutual funds virtually always perform relatively less well, and last year's worst mutual funds virtually always perform relatively better, during the next year.

Information bias, analogously, refers to a tendency to understate or overstate exposure–disease association due to asymmetries in a data collection process after sample selection. Recall bias is a well-known special case. For example, women who have a child with a birth defect are more likely to recall all drug use during their pregnancy than are mothers with a healthy child. Mothers with children without a birth defect may simply have forgotten they took certain medication because their attention has not previously been directed to it, while mothers of affected children are more likely to have gone over in their minds everything that occurred during their pregnancy that might explain their children's injuries. Similarly, patients' quality-of-life assessments may be lower if assessed during physician visits occasioned by exacerbations of their chronic diseases, rather than at random times.

Confounding is the misattribution error committed by a Florida resident who, on being informed that Floridians suffer twice the per capita mortality of Alaskans, moves to Alaska to prolong his lifespan. He has neglected to consider that this effect may be due to Floridians being in aggregate older, for which no move northward can compensate. The potential for confounding is present whenever a third variable associated with both exposure and disease in a study sample is unrecognized in the ensuing analysis, allowing effects of the third variable, the confounder, to masquerade as effects of the exposure. In the preceding example, the exposure is state of residence, the outcome is death, and the confounder is age. Confounding can be controlled in numerous ways, most effectively in principle by basing inferences only on comparisons between subjects who are homogeneous with respect to the confounder(s) or by randomized experimentation. But in observational studies, where randomization is impossible, confounding is an ever-present threat because confounding may lurk in the underlying population and propagate to the sample, or arise, sometimes unobserved or unrecognized, from selection bias, information bias, or nonrandom loss of follow-up data.

Invalid internal inferences can also arise solely due to chance simply because relationships in samples inevitably differ somewhat from those in the underlying populations. Measurements of quantities inevitably vary somewhat around the true values, and categorizations including diagnoses are subject to random misassignments (as distinct from systematic calibration errors).[54] Large departures tend to occur less frequently than small ones, but when they do occur in samples, they can seriously mischaracterize the populations or processes from which

54. Thus, literally projecting relationships observed in samples back to their encompassing populations subjects inferences to random instability. The degree of instability is commensurate with the degree of underlying variability resulting from the combination of population heterogeneity with measurement variability. But it is roughly inversely proportional, on the scale of the measurements studied, to the square root of the number of study participants.

those samples arise. In principle, the chance misfortune of having obtained unrepresentative data is a potential alternative explanation to a systematic causal association for any associations seen in research data. Methods are thus needed for deciding when this ever-present potential alternative explanation is sufficiently implausible to ignore, thus allowing valid internal inference from sample to population.

A variety of methods, based on several statistical schools of thought, have been proposed for doing this. Although this may be changing, the dominant methods in the scientific literature for making such assessments have been *p*-values, confidence intervals, and the statistical-hypothesis tests they embody.

A "*p*-value" is a statistic—that is, a value calculated from a collection of data—that localizes some specific quantitative summary of the data, most simply a mean, proportion, or correlation within a distribution of possible values it might have taken were the research to be repeated many times. Consider the following example.[55]

Suppose researchers wish to study whether drug X improves the health outcomes of men and women equally. The researcher draws a sample of 50 men and 50 women at random from 5,000 males and 5,000 females exposed to a particular medication. These individuals are examined to ascertain whether the medication had improved their health outcome, and the results indicate the drug helped 38% of the men and 58% of the women. The *p*-value answers the following question: If the drug is equally effective for all 5,000 males and 5,000 females, for what fraction of possible random samples of 50 men and 50 women would researchers have found a discrepancy (suitably measured) as large as or larger than that observed in the current sample?

The assumption that the drug's effectiveness in the population is the same for men and women is called the null hypothesis. The null hypothesis asserts that there is no difference between the fractions of men and women in the whole population whose health outcomes the drug will improve and therefore that differences in the fractions between the samples of men and women are thus due to random variation—the luck of the draw.

The *p*-value is the probability of getting data with a male-female discrepancy as extreme as or more extreme than in the actual data, given that the null hypothesis is true:

$p$ = Probability (as or more extreme data | null hypothesis in model)[56]

---

55. The example is adapted from Chapter 5 in FAIGMAN ET AL., *supra* note 11, § 5.34.

56. Since $p$ is calculated by assuming the null hypothesis is correct (no real difference in improvement rates), the *p*-value cannot give the chance that this hypothesis is true. The *p*-value merely gives the chance of getting evidence against the null hypothesis as strong as or stronger than the evidence at hand—assuming the null hypothesis to be correct. No matter how many samples are obtained, the null hypothesis is either always right or always wrong. Chance affects the data, not the hypothesis. With the frequentist interpretation of chance, there is no meaningful way to assign a numerical probability to the null hypothesis.

A small $p$-value supplies some evidence of disparate impact, but a large $p$-value would merely reflect insufficient evidence in the data to confidently refute the null hypothesis. Regrettably, multiple negatives are involved here. A statistical test is essentially an argument by contradiction. The "null hypothesis" asserts no difference in the population—that is, no disparate impact. Small $p$-values speak against the null hypothesis—there is disparate impact because the observed difference is hard to explain by chance alone.[57] Conversely, large $p$-values indicate that the data are compatible with the null hypothesis: the observed difference is easy to explain by chance (though this may be because the data are not very informative, rather than because the null hypothesis is true). In this example, $p = 4.5\%$. If the null hypothesis is true and, as here, 48 total events occur among 5,000 members of each sex, there is only a 4.5% chance of getting a difference in health outcomes of 20 percentage points or more. The $p$-value for the observed discrepancy is 4.5%, or 0.045.

The $p$-value is often thought of as a measure of the plausibility of the null hypothesis in light of the data collected by a study. In practice, statistical analysts often use certain preset "significance levels"—typically 5% or 1%—to aid in interpretation. [58] Thus "significant" is merely a label for certain $p$-values. Historically, when analysts have used the term significant, they mean only that the

---

*See* David H. Kaye, *Statistical Significance and the Burden of Persuasion*, 46 L. & CONTEMP. PROBS. 13, 21–23 (1983).

57. Thus, if the $p$-value is low, this forces the researcher to choose between two competing explanations: either a) the null hypothesis is false; or b) the null hypothesis is true, but nevertheless an event of small probability has occurred. It seems reasonable to regard b) as implausible, and thus a) as the preferable explanation, leading to rejection of the null hypothesis. The smaller the $p$-value, the more confidence one can have in this conclusion.

58. Much smaller values are used in "omics" studies, to account for simultaneous performance of thousands, tens of thousands, or millions of tests. Note that appropriately compensating in some manner for multiple simultaneous tests is of great importance whenever such tests, even just a few, are used to address the same basic scientific question. If each test is conducted at $\alpha = 5\%$, chance immediately becomes a much more plausible alternative explanation for false positive results, vitiating the claim that the testing procedure controls invalid results due to chance variation. This issue commonly arises in longitudinal research, when individuals are studied by repeated measurement of the same characteristics over time and when multiple outcomes are of interest—as is virtually always the case—in studies of health care. In such situations, one needs to choose an $\alpha$ much smaller than 0.05. For example, researchers sifting through a large number of correlations in search of significant relationships will inevitably find some. A 0.05 significance test means that one time in 20 we should expect to observe a relationship this strong even when the null hypothesis is true. An example of this phenomenon may be found in Patricia Shiono and Mark Klebanoff's study concerning the teratogenicity of the drug Bendectin. Patricia Shiono & Mark Klebanoff, *Bendectin and Human Congenital Malformations*, 40 TERATOLOGY 151, 151–55 (1989). The authors separately examined the relationship between the drug and 58 categories of malformation. Bendectin was significantly correlated with 3 types of defects: microcephaly, congenital cataract, and lung malformations. The authors note that the three significant correlations are exactly the number one would expect by chance using a .05 significance test for approximately 60 relationships. In such situations, one needs to choose an $\alpha$ much smaller than 0.05.

$p$-value is below some such conventional threshold value.[59] In principle, publication of a $p$-value allows every reader to apply his or her own personal statistical significance threshold to each particular situation. In the above example, the observed discrepancy of 4.5% is significant at a 5%, but not a 1%, threshold.

The Greek letter $\alpha$ (alpha) is often used to symbolize such a consensus "level," or conventional threshold that, when used consistently, represents the fraction of false positive tests of true null hypotheses deemed acceptable. Since $\alpha$ may be set at any value, this approach is potentially very effective at limiting false positive results due to chance. Rejecting the null hypothesis when it is true is known as a Type I error. However, the inevitable cost of setting a stringent threshold (i.e., low $\alpha$ for rejecting the null hypothesis) is an increase in the risk of making the opposite mistake—false negative error: failing to reject the null hypothesis when in fact there is a relationship. This is called a Type II error. The Greek letter $\beta$ (beta) is often used to symbolize this risk. It depends on the specific alternative hypothesis entertained when the null is rejected. Figure 1 indicates these two ways in which we may draw an incorrect conclusion based on sample data. Here, $H_0$ denotes the null hypothesis and $H_1$ the alternative hypothesis, and thus the Type II error rate $\beta$ depends on the choice of the alternative hypothesis $H_1$.

| | | Population Reality | |
|---|---|---|---|
| | | $H_0$ true | $H_1$ true |
| Decision Based on Sample Data | $D_1$: Reject $H_0$ | False Positive (Type I) Error<br>[$\alpha$ = Probability of Wrongly Rejecting $H_0$ When $H_0$ is True] | True Positive<br>No Error<br>[$1-\beta$ = Probability of Correctly Affirming $H_1$ When True] |
| | $D_0$: Accept $H_0$ | True Negative<br>No Error<br>[$1-\alpha$ = Probability of Correctly Retaining $H_0$ When True] | False Negative (Type II) Error<br>[$\beta$ = Probability of Wrongly Retaining $H_0$ When $H_1$ is True] |

*Figure 1: Possible Decisions and Errors*

The quantity $1-\beta$, which varies between studies even when $\alpha$ is held fixed, is called the "statistical power" of the hypothesis test. When power is sufficiently high, typically 80% or 90%, it is reasonable and conventional for medical clinical trials to regard a nonsignificant test as affirming the null hypothesis or something close to it. Otherwise, a nonsignificant test reflects data simply insufficient in sample size and/or measurement precision, relative to population and measurement variability, to differentiate a true relationship from a truly null situation while

59. Because the $p$-value is affected by sample size, it does not measure the extent or importance of a difference. Suppose in our example we had a sample of 50,000 or even 500,000 of each sex. A small change in health outcomes of two or three percent may be statistically significant but may be of little practical significance.

maintaining the stipulated false positive rate, $\alpha$.[60] In these circumstances the null hypothesis is "retained," not accepted, but remains a plausible alternative explanation for any observed relationship, thus invalidating an inferential claim based on this relationship.

Statistical significance testing, as described above, may be applied not just to null hypotheses but also to hypotheses positing non-null statistical relationships of any specific strength represented by an association measure such as a correlation or regression coefficient, or a ratio of rates (equivalently, "hazards"), risks, or odds. The set of all values measuring the strength of a relationship that are retained by a statistical significance test at level $\alpha$ constitutes a "$100(1-\alpha)$% confidence interval." Such a confidence interval can be interpreted as the range of plausible strengths of association (in the underlying population or otherwise characterizing the data-generation process) that are compatible with the data collected by a study.

The paradigm described above, although frequently used in scientific studies, is not the only way to assess the possible role of random chance in research. Many prominent statisticians and other scientists have questioned it, and the need for change is increasingly accepted.[61] Alternatives to such classical-statistical-frequentist significance testing based on quite different rationales, such as Bayesian[62] and empirical Bayesian inference, are ascendant and are likely to become

---

60. Power is a function of a study's sample size, the variability of contributing measurements, the size of the effect one would like to be able to detect—i.e., the alternative hypothesis—and the significance level used to guard against Type I error. Because power is a function of the significance level, all other things being equal, minimizing the probability of one type of error can be done only by increasing the probability of the other. For a useful technical discussion of study designs given the relative importance of avoiding Type I or Type II errors, see Michelle Burtis et al., *Error Costs, Legal Standards of Proof, and Statistical Significance*, 25 SUP. CT. ECON. REV. 1 (2017).

Because the power of any test is reduced as the size of an effect decreases, Type II threats to causal conclusions are particularly relevant with respect to rare events. Surprisingly, relatively few legal cases contain power discussions. *See* Merrell Dow Pharms., Inc. v. Havner, 953 S.W.2d 706, 722–23 (Tex. 1997); Miller v. Pfizer, Inc., 196 F. Supp. 2d 1062, 1074 (D. Kan. 2002), *aff'd*, 356 F.3d 1326 (10th Cir. 2004); Doe 93 v. Sec'y of Health & Hum. Servs., 98 Fed. Cl. 553, 568 (2011); Jon T. Powell, *How to Tell the Truth with Statistics: A New Statistical Approach to Analyzing the Bendectin Epidemiological Data in the Aftermath of* Daubert v. Merrell Dow Pharmaceuticals, 31 HOUS. L. REV. 1241, 1266–67 (1994) (setting forth formulae for calculating the power of case-control and cohort studies from 2 x 2 contingency table data).

61. Ronald L. Wasserstein and Nicole A. Lazar, *The ASA Statement on p-Values: Context, Process, and Purpose*, 72(2) AMER. STATISTICIAN 129 (2016).

62. The Bayesian approach, unlike the frequentist one, regards unknown quantities ("parameters") about which inference is required—in the above example, the true success rates in the 5,000 men and 5,000 women—as themselves having probability distributions—interpreted, however, as representing epistemic uncertainty rather than physical randomness. One starts with a "prior distribution" for these quantities, representing uncertainty ahead of seeing the data. This uncertainty is typically high and so represented by a diffuse prior distribution, but genuine prior knowledge, perhaps from other studies, can be incorporated. There is a formal method based on Bayes's theorem (a mathematical result in probability theory) for computing the "posterior distribution," which updates the prior using

more prominent in the scientific literature over time.[63] But the preceding methods are deeply rooted in current scientific practice. When used properly, they can help discriminate between research observations for which chance is a more or less plausible explanation, and thus between research findings for which chance is a stronger or weaker threat to research validity. They are thus likely to continue to be relevant to legal work for some time.

*2. External Validity*

Internal validity of a finding refers to the strength of its support, uniquely among other explanations, among groups of subjects and under conditions like those of the studies from which the finding was derived. External validity, in contrast, refers to the rationale for extrapolating the finding to substantially different subjects and/or conditions.[64] If the finding, as is typical, consists of a measure of strength of a relationship, then the finding is externally valid if that relationship persists and remains similarly strong, despite changes in the nature of the subjects studied, the conditions under which they are tested, and the ways in which the related variables are measured.

Thus, if a study uncovers an apparent cause-and-effect relationship, the researcher must determine to which categories of individuals the relationship can be generalized. For example, if a study includes only adult men as subjects, the researcher must determine whether the results can be generalized to women. When there are alterations in the relationship with changes in the characteristics of subjects, particular conditions, or measurement processes, then the variable with which the relationship changes is called an effect modifier and is said to statistically "interact" with the exposure in predicting the outcome. For example, if women who take Drug A tend to live longer than otherwise, but men taking Drug A tend to die earlier, then gender is said to modify the relationship of Drug A to mortality, and Drug A and gender are said to interact in predicting death. Potential interactions

---

the information in the data, so representing the final state of knowledge of the unknown parameters. A Bayesian analogue of a *p*-value, in the example, could be the posterior probability that the true success rate for women exceeds that for men. In some cases, and for specific prior distributions, such a Bayesian measure can be numerically the same as, or close to, a frequentist *p*-value; but its interpretation would be entirely different. There is also a Bayesian analogue of a confidence interval, the "credible interval," which has a chosen posterior probability of including the targeted unknown parameter. Again there can sometimes be numerical similarities, but important interpretive distinctions, between frequentist confidence intervals and Bayesian credible intervals. However, because the fundamental output of a Bayesian analysis is a full posterior probability distribution for all unknown parameters, this can be applied in much more flexible ways, not mimicking any classical approach. ANDREW GELMAN ET AL., BAYESIAN DATA ANALYSIS (3d ed. 2014). For a law review article espousing this approach, see Neal C. Stout & Peter A. Valberg, *Bayes' Law, Sequential Uncertainties, and Evidence of Causation in Toxic Tort Cases*, 38 U. MICH. J.L. REFORM 781 (2005).

63. For useful discussions of alternative approaches, see ANTHONY WILLIAM FAIRBANK EDWARDS, LIKELIHOOD (1992); GELMAN ET AL., *supra* note 62. For a case in which an expert uses a Bayesian approach, see *Milward v. Acuity Specialty Products Group, Inc.*, 969 F. Supp. 2d 101, 105–06 (D. Mass. 2013).

64. For a discussion of external validity threats, see FAIGMAN ET AL., *supra* note 11, § 4:39.

between factors—such as the one between subjects and treatment in this example—compromise external validity.[65]

Other examples involve the ability to generalize across race, ethnicity, and class. Similar considerations apply to conditions under which an effect is said to occur. Thrombolytic agents such as tissue plasminogen activator (tPA) are highly beneficial when administered within a few hours after onset of ischemic stroke, when the prospect of lysing the blot clot and thereby sparing brain tissue from oxygen starvation and cell death outweighs the chance of initiating or exacerbating intracranial bleeding, which itself can cause death of cells otherwise undamaged by the initial stroke. However, hours later the chance of saving tissue has passed, while the chance of bleeding has not, and so tPA use then is clearly disadvantageous.[66] Time of administration since stroke onset thus modifies the relationship of antithrombolytic therapy to death and disability after ischemic stroke.

External validity is of evident major concern in considering the implications of animal studies and *in vitro* research for human health, because interspecies extrapolations and extrapolations across vast dose ranges are frequently involved. Interaction of research setting with treatment is another threat to external validity. Even well-crafted experiments that do their best to assure internal validity cannot necessarily ensure transferability from the laboratory to the world at large or even, in some cases, to other laboratories. In the arena of toxic torts, this is often the case because levels of exposure in a laboratory are not the same as exposures in the world at large, an issue to which we return below.[67] In behavioral studies of humans, ethical considerations greatly limit the verisimilitude of decision-making, given the artificiality of the laboratory setting. This is often a major consideration in scientific evaluations of research in the forensic sciences.

### *B. Types of Evidence Available to Establish Causation*

Courts almost universally agree that one must "rule in" the putative cause of an injury before ruling out alternative causes.[68] The goal is to persuade a court that a substance is capable of causing the illness under consideration. Efforts to answer this general causation question typically rely on several bodies of scientific knowledge. These include toxicological, epidemiological, and other group-based data on humans as well as genetic information.

---

65. This example highlights an important point. One primary reason we need to be concerned about external validity is that altering the group in question (women instead of men) introduces a potential effect modifier that may alter the relationship between two other variables (drug usage and life expectancy).

66. Edward C. Jauch et al., *Guidelines for the Early Management of Patients with Acute Ischemic Stroke: A Guideline for Healthcare Professionals from the American Heart Association/American Stroke Association*, 44 STROKE 870 (2013).

67. *See infra* Part III.B.1.

68. For a review of the many cases adopting this position, see FAIGMAN ET AL., *supra* note 11, § 21:6.

*1. Toxicology*[69]

Toxicological science includes research at several different levels: molecules, cells and other biological materials, nonhuman animals, and human beings. Mechanistic toxicology studies the cellular, biochemical, and molecular basis by which substances affect biological targets. Part of this endeavor involves examining a substance's chemical structure and comparing it to other compounds of similar structure whose biological activity is better understood and for which there is existing toxicity information. Thus, the process sometimes goes by the name "structure-activity analysis." Structure-activity relationships can expedite the identification of potentially beneficial or harmful substances. However, modest differences in chemical structure can lead to different levels of toxicity, in part because of the way organisms metabolize the substance. Therefore, a structure-activity analysis is not determinative of toxic effects.[70]

*In vitro* testing provides an example of this principle. *In vitro* testing is conducted on bacteria, human or animal cells, isolated tissues, embryos, or organs. Such testing takes place outside a living organism. Cultures in a test tube or petri dish are exposed to potential toxicants (or drugs). Toxicologists study the perfusion of the substance through the culture and assess biological responses. These methods are useful to mechanistic toxicologists because they can provide insight to the mechanisms of toxicity, such as specifically how potentially cancer-causing substances may damage DNA or cause other changes in a cell nucleus.

There are advantages to *in vitro* testing. It is much cheaper than tests on living organisms. Additionally, living organisms are very complex, so it is often difficult to identify interactions and processes of interest in humans. *In vitro* systems are simpler than intact organisms, making it easier to isolate and study biological interactions such as the ways immune system proteins attach themselves to antigens.

However, the most important shortcoming of *in vitro* tests from the law's perspective is their low external validity. The limited ability to extrapolate *in vitro* test findings to effects of the substance on living organisms is a direct consequence of *in vitro* systems' simplicity because they cannot account fully for the environment in which events occur within the organism. With respect to potential new pharmacotherapies, for example, effectiveness *in vitro* more often than not fails to translate into effectiveness in living animals. Impediments to extrapolation include the following factors: a chemical may not be absorbed by living organisms; the chemical may be distributed in a living organism such that more (or less) reaches specific locations than would be predicted based on its absorption; and the chemical

---

69. For an overview of this discipline by Joseph Rodricks, see *id.* § 22.17–.33.

70. For example, benzene and the alkyl benzenes, e.g., toluene, xylene and ethylbenzene, share a similar structure. This similarity is reflected in the fact that acute exposure to each of them produces similar central nervous system anesthetic-like effects. However, only benzene causes leukemia and damage to bone marrow. The damage is caused not by benzene itself but a toxic metabolite of benzene. Cliona M. McHale et al., *Current Understanding of the Mechanism of Benzene-Induced Leukemia in Humans: Implications for Risk Assessment*, 32(2) CARCINOGENESIS 240 (2012).

may be rapidly metabolized into a form or substance with a different activity profile than the parent agent.[71]

*In vivo* testing, by contrast, is performed on living organisms other than humans. Toxicologists investigate the life cycle of substances in the body: how they are absorbed, distributed, metabolized, and eliminated from the organism. Such research may be directly designed to study the impact of a substance on the test species itself, but frequently its purpose is to tell us something about the substance's potential impact on humans.[72]

The primary strength of well-conducted animal studies is the ability to exercise strong control of extraneous factors and thus assure internal validity. Test animals may be inbred to be quite homogeneous and randomly assigned to treatment or control so that, with a sufficient sample size, we can be reasonably certain that any observed effect is due to the treatment. Moreover, the best studies are blinded so that the researcher doing the experiment does not know whether a given animal received a treatment or a placebo. A blinded study with randomized assignment to treatment and sample size providing adequate power has very high internal validity because, by design, it removes or substantially mitigates most forms of bias. In these circumstances, we can be reasonably certain the exposure is the primary cause of observed average differences in outcome, if such differences exceed what can be expected from chance variation.[73]

However, extrapolation across species inevitably raises questions of external validity. Thalidomide, a catastrophic teratogen in humans, provides an instructive example. When it was initially used as a remedy for nausea in pregnant women, it caused fetal limb malformations (phocomelia). Earlier *in vivo* testing using rats had failed to indicate any malformations. Only subsequently did we come to understand that while the drug produces similarly severe results in humans, rabbits, and monkeys, this adverse consequence does not seem to occur in the rats

---

71. *See* FAIGMAN ET AL., *supra* note 11, § 22.21.

72. The choice of which animal models to employ is a complex one, influenced by the end point of interest (e.g., cancer, birth defects, etc.), similarities between the animal and human systems of interest, and, at a practical level, the costs involved in testing on different species. The design of an animal study will vary depending on the type of injury—acute, chronic, reproductive—one is investigating. *See id.* § 22.22.

73. Unfortunately, a fair percentage of animal studies fail to conform to this ideal and their results are appropriately suspect. For example, one study evaluated 290 animal study abstracts containing two or more experimental groups that were accepted by the Society for Academic Emergency Medicine. Vik Bebarta et al., *Emergency Medicine Animal Research: Does Use of Randomization and Blinding Affect the Results?*, 10 ACAD. EMERG. MED. 684, 684–87 (2003). One hundred and ninety-four were not randomized, and 259 were not blinded. The non-randomized and non-blinded studies had 3.4- and 3.2-fold higher odds, respectively, of claiming a statistically significant outcome than did those that were randomized and blinded. *Id.*; *see also* Nicolas A. Crossley et al., *Empirical Evidence of Bias in the Design of Experimental Stroke Studies: A Metaepidemiologic Approach*, 39 STROKE 929, 929–34 (2008); John P.A. Ioannidis, *Extrapolating from Animals to Humans*, 4 SCI. TRANSLATIONAL MED. 151 (2012). Before admitting expert testimony based on an animal study, it is highly advisable to assess whether the study randomly assigned animals to treatments and used a blinded design.

in which it had been tested, nor in mice, hamsters, or other rodent species commonly used in toxicological testing.[74]

Species differences are not the only threat to external validity posed by animal studies. Dosage is also an issue. In some studies, the goal is to assess the acute toxicity of a chemical, that is, the results from a single dose of the substance. In other studies, the goal is to assess the toxicity of repeated or continuous exposure over a longer or shorter time period. With respect to acute response studies, animal researchers traditionally employed a quantal-response experiment to determine the $LD_{50}$ (i.e., the dosage, usually measured as milligrams per kilogram weight of the animal, needed to kill half the animals in a study within a relatively short period of time).[75] Costs and evolving sensibilities regarding treatment of animals have caused researchers to use this design less frequently.[76] In studies not focused on acute poisoning, where the question is whether the drug causes cancer or similar adverse outcomes, animals may be given the maximum tolerated dose ("MTD"), which is defined conceptually as a dose just below what may cause premature mortality due to short-term toxic effects.[77]

Why use such high doses? Many substances produce an adverse effect in only a small percentage of organisms when ingested at a dose similar to that encountered in the environment. Given this reality, at lower doses it would take a prohibitively large number of animal subjects to detect a substance's adverse effects with any reliability. Smaller samples would generate an unacceptably large number of Type II (i.e., false negative inferential errors, the failure to detect a toxic effect when it exists).[78] Consequently, given the expense of animal studies, researchers assessing whether some substance is toxic may expose animals to relatively large doses to ascertain if there is any effect (and to guard against the potential for false negative results). If there is a positive result, toxicologists must then extrapolate a predicted incidence at a more realistic lower dose rate.

---

74. Max Sherman & Steven Strauss, *Thalidomide: A Twenty-Five Year Perspective*, 41 FOOD DRUG & COSM. L.J. 458, 464 (1986).

75. A substance with a median lethal dose of 1 mg/kg is generally considered highly toxic. A substance with a median lethal dose of greater that 500 mg/kg is considered slightly toxic.

76. *See* FAIGMAN ET AL., *supra* note 11, § 22.23.

77. *See* Bernard D. Goldstein & Mary Sue Henifin, *Reference Guide on Toxicology*, *in* REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 401, 408 (2d ed. 2000).

78. For example, if an adverse outcome occurs naturally—i.e., in the absence of the exposure under investigation—in 2 of 100 cases, and exposure to the toxin at approximately the dose humans would experience increases the incidence of the adverse outcome to 3 in 100 (a fifty percent increase), a study using 200 animals (100 exposed and 100 controls) that resulted this outcome would not be statistically significant (at the 5% level). Indeed, if these findings were replicated using a thousand animals (i.e., 15 of 500 sick in the exposed group and 10 of 500 in the control group), the results would still not be statistically significant. If, on the other hand, we increased the dose to ten times what humans were likely to be exposed to, and at this greater dose 8 in 100 animals suffered the adverse effect under investigation, a study producing this result with 200 animals would reach statistical significance. And even this example is misleading because actual effects for animals and humans may not be 3 in 100 but rather 3 in 1,000, or even less frequent.

Extrapolation involves several issues. The first is simply the question of how to equate a dose given an animal to a similar human dose. If one gives a laboratory animal an X mg dose, what is the human equivalent? The question falls within the field of allometry: the study of biological size and its consequences. Unfortunately, there is apparently no uniformly accepted formula for such inter-species extrapolations. Traditionally, three methods of extrapolation have been used: body-mass equivalence, surface-area equivalence, and caloric scaling across species.[79] However, advances in pharmacokinetics (the study of the absorption, distribution, metabolism, and excretion of drugs over time after introduction) and pharmacodynamics (the study of the time course of the therapeutic and other drug effects in relation to the dose and route of administration) [80] suggest other considerations as well, such as differing bioavailabilities[81] of a substance across species. Once one has resolved the dose-equivalence relationship, the basic issue of extrapolation from high to low doses remains quite difficult, and estimating low-dose effects with high precision has not yet proved, and may never prove, to be possible. There are a number of ways in which high-dose toxicity testing differs from lower dose effects: there may be limits to the solubility of the compound; enzymes may become saturated at high doses, limiting absorption; detoxification mechanisms in the liver and elsewhere may be saturated; and metabolites may cause toxicity that would not occur with lower doses. Each of these factors may produce nonlinear effects, making extrapolation to dosages that reflect typical human exposure problematic. This is especially the case if the animals were only exposed to the MTD.[82]

Finally, we should note that animal studies may fail to capture human experience because the nature of the exposure and/or injury differs from their human counterparts. That was the problem with an animal study discussed in *General Electric v. Joiner*.[83] The plaintiff claimed that exposure to PCBs had advanced the time of the onset of his lung cancer. The infant mice in the study in question were

---

79. K. Schneider et al., *Allometric Principles for Interspecies Extrapolation in Toxicological Risk Assessment—Empirical Investigations*, 39(3) REGUL. TOXICOLOGY & PHARMACOLOGY 334 (2004).

80. Pharmacokinetics is a branch of pharmacology devoted to understanding what happens to substances entering a living organism. These include any chemical xenobiotic such as drugs, pesticides, and other chemicals.

81. Bioavailability is a subcategory of absorption and is the fraction of an administered dose of a substance that reaches the systemic circulation in the body. It is one of the principal pharmacokinetic properties of substances. BASIC PHARMACOKINETICS AND PHARMACODYNAMICS: AN INTEGRATED TEXTBOOK AND COMPUTER SIMULATIONS 44–45 (Sarah Rosenbaum ed., 2d ed. 2017).

82. When regulatory agencies need to set a permissible dose based on animal study research, they traditionally conduct tests designed to uncover a "no observed adverse effect level" (NOAEL), which is defined as the highest dose at which there is no statistically significant difference between subject animals and controls. Increasingly, agencies use a slightly different approach called a Benchmark Dose (BMD). The agency objective is to establish a reference dose (RfD) or reference concentration (RfC) of a toxin—i.e., a dose or a concentration "that is likely to be without an appreciable risk of deleterious effects during a lifetime." *See* FAIGMAN ET AL., *supra* note 11, § 22.24.

83. 522 U.S. 136, 144 (1997).

exposed in a different way and exhibited a different injury than that suffered by the plaintiff. Therefore, the study–plaintiff relationship was untenable. This problem occurs in other cases as well.[84]

Many of the issues discussed above also apply to human toxicology. Specifically, the study of pharmacokinetics and pharmacodynamics of substances in humans plays an increasingly important role in assessing the biological plausibility of a hypothesis that a substance may cause injury to individuals. An important part of this inquiry involves estimates of the dose to which an individual is exposed. The "first law" of toxicology is that the dose makes the poison.[85] With respect to substances that are thought to have some threshold before they produce adverse effects, ascertaining dose is key to determining if a particular injury is reasonably attributable to a given substance. Toxicology can provide information as to the threshold at which a substance becomes harmful.[86] Unfortunately, in many tort situations, dosage is very hard to ascertain. And to complicate matters, dose per se is not the only consideration. One must also understand the bioavailability of the substance and attributes such as the age of the exposed individual.

*2. Epidemiology*[87]

Epidemiologists search for and assess potentially causal exposure–disease relationships in observational data using two contrasting general research strategies: cohort studies and case-control studies. Cohort studies compare the incidence of an injury in groups of persons exposed to a substance to the incidence in groups of persons not exposed or between groups otherwise varying in their levels of exposure. Case-control studies, on the other hand, compare the degrees of preceding exposure experienced by groups of persons who do ("cases") or do not ("controls") have an illness or injury under investigation or among groups whose members differ in the severity of illness or injury. Cohort or case-control studies are further designated as "prospective" or "retrospective" depending on whether the data analyzed were collected using methods specifically planned in advance (prospective) for this or a closely related purpose, or whether they were collected previously for other purposes and are being analyzed later (retrospective). The "prospective vs. retrospective" classification is often a crude proxy for data quality for the purposes of assessing a specific investigation. Epidemiological literature can be confusing in this regard because the respective terms "prospective" and "retrospective" were historically used as virtual synonyms for cohort and case control research. However,

---

84. *See* McCarty v. Arch Wood Prot., Inc., No. 11-109-HRW_CJS, 2016 WL 2936435, at *11 (E.D. Ky. Feb. 26, 2016*), report and recommendation adopted*, No. CV 11-109-HRW, 2016 WL 1306067 (E.D. Ky. Mar. 31, 2016); *In re* Mirena IUD Prods. Liab. Litig., 169 F. Supp. 3d 396 (S.D.N.Y. 2016).

85. The second law concerns the specificity of toxic effects of individual chemicals. Due to the unique chemical structure of the agent and the laws of biology that govern the response, chemicals cause some specific injuries. The third law is that humans are animals and, therefore, the study of animals can provide useful insight into effects in humans. *See* Goldstein & Henifin, *supra* note 77, at 401.

86. Some substances are thought to have no threshold below which they are not harmful to some degree. Ionizing radiation is one example.

87. *See* FAIGMAN ET AL., *supra* note 11, § 23.30–.47 (overview of this discipline by Noel Weiss).

that former usage has been replaced by their current characterization of the temporal relationship between the investigation and the data collection process.

Depending on the disease context and study design, several related statistics are most commonly used by epidemiologists to express the strength of a relationship between exposure and the occurrence of an injury. In cohort studies, when the time interval between the exposure and recognition of the injury is of little relevance, either because these events are in close proximity such as in a gastroenteritis outbreak after a contaminated meal, or the intervals between them are otherwise quite uniform or observed only over a fixed follow-up interval such as a year, the proportion experiencing the injury within a group is termed the "risk," and the ratio of the risks in groups differing in exposure is termed the "relative risk (RR)" or synonymously the "risk ratio." When the time interval varies substantially in a way that matters, as is typically the case in studies of chronic diseases such as cardiovascular disease and cancer, the "hazard ratio (HR)" is a superior measure of the strength of the underlying etiological process. The hazard ratio is a ratio of the rates at which disease events occur over time relative to the gradually diminishing sizes of the remaining unaffected groups.

In case-control studies, the "odds ratio (OR)" is used. The odds ratio represents the odds that an exposure has occurred given a particular outcome, compared to the odds of the exposure having occurred in the absence of that outcome. For events that rarely occur, or at least are quite uncommon during a time period studied, the numerical values of these three association measures in the same population will be quite similar, and the odds ratio from a sample is often used to approximate a risk ratio and/or hazard ratio, leading to some slurring of the technical terminology in the scientific literature.

Epidemiologists also use additional measures, known as "attributable risks" and either "attributable fractions" or "etiologic fractions," to assess the aggregate impact of an exposure on disease occurrence in an exposed group and in mixed populations of exposed and unexposed persons. Of particular relevance in the legal context is the attributable fraction among the exposed = $1 - 1 / RR$, which exceeds one half when $RR > 2$, providing the "more likely than not" preponderance of evidence basis for using $RR > 2$ as a justification for awards in many tort cases.[88]

---

88. For example, imagine that in a certain area there were 5,000 newborns and the mothers of 1,000 of these children took Bendectin during pregnancy. Among the 1,000 children whose mothers took Bendectin, 60 were born with and 940 born without a specific birth defect. Presuming the defects become manifest in utero and are observable at birth, the time dimension is constant and may be disregarded. Among the 4,000 children whose mothers did not take Bendectin, we find that 160 were born with and 3,840 were born without that defect. The cumulative incidence, also known as the risk, among the exposed children is $60 / (60 + 940) = .06$ and among nonexposed children is $160 / (160 + 3840) = .04$. The relative risk is thus $.06 / .04 = 1.50$. Among the $60 + 160 = 220$ cases born with the defect, the odds of the mother having taken Benedictin were $60 / 160 = 0.375$, while the corresponding odds

Epidemiological studies have several advantages over other toxicological studies. They share with animal studies the advantage that they measure the effect of a substance on a whole organism, not simply the effect on a cell culture or an organ. And they sidestep the difficult cross-species comparisons confronting animal studies when those studies are used to predict effects on humans. In sum, they confront fewer external validity challenges to their results.

However, cohort and case-control studies have disadvantages as well. Because they are observational studies rather than true experiments, there is always the possibility that the true relationship between a "cause" and an "effect" has been distorted by failure to account for some unmeasured confounder(s) linked to both. This problem is one of internal validity—our confidence that the study design warrants a conclusion that an observed correlation reflects a causal relationship rather than one due to confounding or another source of methodological deficiency.

Noel Weiss provides the following example. In a study to investigate the effect of fertility-drug use on ovarian-tumor incidence, the exposed group is women who took certain fertility drugs. The comparison, however, cannot simply be women who did not take the drugs, because some in such a group would have borne children, which itself reduces the chances of developing ovarian cancer. A proper comparison group, therefore, is comprised of other infertile women. Using such a control group avoids a potential selection threat to the validity of the study. Failure to do so risks erroneously inferring that fertility drugs increase ovarian cancer solely because fertile women do not need infertility drugs.[89] As we noted above, there are several well-understood general threats to internal validity of observational studies, including reverse causation, selection bias, measurement bias, and confounding.[90]

*3. Clinical Trials*

Clinical trials have relatively higher external validity than animal studies because they do not require cross-species extrapolations. When properly done, they

---

among mothers of those born without the defect were 940 / 3840 = 0.245, yielding an odds ratio of 1.53, very similar to the relative risk of 1.50 because the birth defect was quite uncommon in both groups.

The attributable fraction among the exposed is 1 − 1/RR = 1 − 1/1.50 = 1/3. Also, the "attributable risk" is the absolute difference between risks of exposed and unexposed groups. In our example, the unexposed risk is .04 and the exposed risk is .06. Thus, the risk attributable to Bendectin, *if* Bendectin use were the only difference between exposed and unexposed children, is .02. The inverse of this attributable risk, 50, is called the "number needed to harm (NNH)," i.e., the number of women taking Bendectin during pregnancy from whose pregnancies one would expect, on average, one child with the particular birth defect. In a mixed population in which 10% of mothers took Benedictin during pregnancy, the "population attributable risk" of the birth defect from Benedictin would be 10% of the attributable risk among the exposed, or 0.10 × 0.02 = 0.002 = 0.2%. Finally, the "population attributable fraction" is the ratio of the population attributable risk to the total risk in the population. Thus, if the birth defect were known to be present in 0.5% percent of newborns in the population, then the population attributable fraction due to Benedictin would be 0.002/0.005 = 40%.

89. Noel Weiss, *General Concepts in Epidemiology*, in FAIGMAN ET AL., *supra* note 11, §23.35.

90. See *supra* pp. 14–15.

also share with animal studies the internal-validity virtues of other true experiments. Imagine, for example, that there is some genetic trait that strongly influences how someone responds to a drug. In a sufficiently large, randomized trial, those who possess this trait are highly likely to be very nearly equally represented in the control (receive the placebo) and experimental (receive the drug) groups, and therefore this trait should not affect the observed relationship between the presence or absence of the drug and the course of the illness the drug is designed to treat. This is equally true regardless of whether the genetic trait is known or unknown to the researchers.

Because of this and other attributes, well-designed clinical trials are sometimes called the "gold standard" of research on toxic injuries. However, some caveats are in order. First, just as was the case in animal studies, a well-designed study should wherever possible employ a blinded or "masked" methodology to avoid measurement bias. In clinical trials, this means that optimally both the patient receiving a drug and the physicians or other health professionals administering the drug and/or assessing its toxic effects are unaware of whether the individual is receiving the drug or its placebo or other comparator. Unfortunately, not all studies do or even can follow such procedures. When they do not, internal validity may be compromised because we cannot be certain whether the observed result is due to the substance in question or the expectations of the patient and/or the physician. A second limitation arises from the fact that cost considerations restrict the size of even large clinical trials. Consequently, they may not detect rare adverse outcomes. Moreover, for a combination of practical and ethical reasons, clinical trials are only rarely performed to assess well-established clinical practices, even when these are not well founded in evidence. Neither can clinical trials normally include some vulnerable groups such as young children. Finally, with few exceptions randomized-trial participants must be restricted to those who can ethically be randomized to any treatment group, and the extent of experimental control and the intensity of clinical follow-up may further restrict participation, making many clinical-trial results most directly relevant to relatively narrow groups of patients.

*4. Susceptibility and Genetic Information*[91]

Individuals vary greatly in their responses to various substances. For example, the relative risk of getting lung cancer is quite high for smokers.[92] Nevertheless, the great majority of average smokers (perhaps 85% to 90%) do not

91. Valuable discussions of the increasing role of genetic information in toxic tort litigation may be found in the following: Russellyn S. Carruth & Bernard D. Goldstein, *Relative Risk Greater than Two in Proof of Causation in Toxic Tort Litigation*, 41 JURIMETRICS J. 195 (2001); Susan R. Poulter, *Genetic Testing in Toxic Injury Litigation: The Path to Scientific Certainty or Blind Alley?*, 41 JURIMETRICS J. 211 (2001); Gary E. Marchant, *Genetic Data in Toxic Tort Litigation*, 14 J.L. & POL'Y 7 (2006); Jamie A. Grodsky, *Genomics and Toxic Torts: Dismantling the Risk-Injury Divide*, 59 STAN. L. REV. 1671 (2007); Steve C. Gold, *The More We Know, The Less Intelligent We Are?– How Genomic Information Should, & Should Not, Change Toxic Tort Causation Doctrine*, 34 HARV. ENVTL. L. REV. 369 (2010).

92. *See* FAIGMAN ET AL., *supra* note 11, § 25.29.

contract this disease.[93] A full understanding of variations in susceptibility is still out of our reach, but some factors are understood.

Increasingly, genetic research is uncovering effects produced by genetic variations. [94] For example, conventional epidemiologic research failed to demonstrate a link between breast cancer and smoking although tobacco smoke contains known mammary carcinogens.[95] As Gold et al. note:

> Genomic investigations observed that variations in the *NAT2* gene, which codes for a carcinogen-neutralizing enzyme, [96] dramatically influenced the breast cancer danger from smoking. Women whose genes coded for the most protective form of the enzyme had no increased risk of breast cancer even if they smoked, but women smokers with less protective forms of the gene were eight times more likely to get breast cancer than were women with the same genotype who did not smoke. It remains true, however, that not all women smokers with the less protective genotype will develop breast cancer, some women smokers develop breast cancer even though they do not have that genotype, and some women develop breast cancer even though they neither smoke nor have that genotype.[97]

In this example, the genetic makeup of the individual alters susceptibility to a toxin.[98] In epidemiological terms, the NAT2 gene modifies the association between smoking and breast cancer, and while smoking causes breast cancer in some women, smoking is neither a necessary nor a sufficient cause in all cases. In some cases, such as sickle-cell disease, one's genetic makeup may cause an injury directly

---

93. This observation underlines the point that a large relative risk does not mean that there is a substantial absolute risk associated with exposure. Too often, reports of increased relative risk fail to clarify that the absolute increase in risk may be very small. *See* Alexandra L.J. Freeman & David J. Spiegelhalter, *Communicating Health Risks in Science Publications: Time for Everyone to Take Responsibility*, 16 BMC MED. 207 (2018).

94. Toxicogenomics is the area of study most directly associated with this type of investigation. It uses *in vivo* and *in vitro* research techniques to research the effects of exposing variations of many genes to a suspected toxin. Changes may involve how an exposure causes a genomic change as well as whether it effects a change in a gene's expression. Epigenetics is the study of genetic expression. "Epigenetics examines the biochemical modifications placed 'above' or 'on top of' DNA, which do not alter the actual sequence of the DNA but can cause a gene to be turned on or off or to be expressed more strongly or weakly." Susan E. Brice & Whitney V. Christian, *The Use of Genetic Evidence to Defend Against Toxic Tort Claims – Part III*, 29 INTELL. PROP. & TECH. L.J. 3, 4 (2017).

95. OFF. ON SMOKING & HEALTH, U.S. DEPT. HEALTH & HUM. SERVS., THE HEALTH CONSEQUENCES OF SMOKING: A REPORT OF THE SURGEON GENERAL (2004).

96. An enzyme is a protein that catalyzes a chemical reaction.

97. STEVE C. GOLD ET AL., NAT'L ACADS. OF SCI., ENG'G & MED., SCIENTIFIC EVIDENCE OF FACTUAL CAUSATION: AN EDUCATIONAL MODULE 168 (2016). This example understates the complexity of the issue of gene variation and susceptibility. Typically, susceptibility variation is the result of multiple genes.

98. *See* Krik v. Schaeffler Grp. USA, Inc., 887 F.3d 376, 382 (8th Cir. 2018).

without the introduction of an external substance.[99] In such situations a genetic explanation may compete with other causal hypotheses.

Biomarkers of exposure offer another important way in which genetic information may affect a causal analysis.[100] In this context, a biomarker is a genetic or other cellular change that occurs when one has been exposed to a substance. We discuss biomarkers more thoroughly in the specific-causation section.

### *C. Putting the Pieces Together: Bradford Hill Indicia*[101]

This last Section of Part III uses the Bradford Hill indicia to assess whether the available evidence supports a general causal conclusion.[102] The discussion is intended to tie together the first two Sections of this Part by indicating how the types of evidence discussed above address or fail to address internal- and external-validity threats that argue against a causal conclusion. The indicia give names to many validity concerns discussed above.

As we noted above, Bradford Hill developed these indicia as a way to assess epidemiological results.[103] Where there is statistical evidence of an elevated risk of a disease among those exposed to a substance, the indicia were intended to aid in assessing whether a connection was causal.[104] That is, Bradford Hill thought of the indicia as a helpful way for assessing general causation in one particular area of research: epidemiology. In this Article, we use the guidelines to assess causation evidence arising from the multiple strands of investigation discussed above. The quality and quantity of data on general causation often directly influence the ability

---

99. Frédéric B. Piel, Martin H. Steinberg & David C. Rees, *Sickle Cell Disease*, 376 NEW ENG. J. MED. 1561 (2017).

100. For example, an important chemical in cigarette smoke, Benzo[a]pyrene (BaP), has been shown by *in vitro* studies to damage the *p53* "tumor suppressor" gene that senses if a cell is dividing uncontrollably and tells the cell to die. This laboratory evidence was later confirmed in people in an epidemiological study, which concluded that "p53 mutations in lung cancer from smokers carry highly significant fingerprints of exposure to tobacco components, especially BaP." *See* FAIGMAN ET AL., *supra* note 11, § 25.29.

101. The legal literature contains several versions of the Bradford Hill list. *See* Michael D. Green et al., *Reference Guide on Epidemiology*, *in* REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 600 (3d ed. 2011).

102. Hill, *supra* note 15. References to the criteria appear in over 120 legal opinions on causation. Some references build a substantial part of their argument around the criteria. *See* Soldo v. Sandoz Pharms. Corp., 244 F. Supp. 2d 434, 473–74 (W.D. Pa. 2003); Gannon v. United States, 571 F. Supp. 2d 615, 624 (E.D. Pa. 2007).

103. Some courts prefer to restrict their use to the assessment of epidemiology for fear that otherwise they will be used as a multifactor test employed to support predetermined conclusions. *See In re* Mirena IUS Levonorgestrel-Related Prods. Liab. Litig. (No. II), 387 F. Supp. 3d 323, 347–48 (S.D.N.Y. 2019). This is a legitimate concern, and we do not mean to suggest that all one needs to do to achieve admissibility is to invoke one or more of these indicia. For a useful discussion of this point, see *In re Viagra (Sildenafil Citrate) and Cialis (Tadalafil) Products Liability Litigation*, 424 F. Supp. 3d 781, 797–99 (N.D. Cal. 2020) (Bradford Hill criteria must be applied reliably).

104. Haack, *supra* note 46, at 275; Frank Woodside & Allison Davis, *The Bradford Hill Criteria: The Forgotten Predicate*, 35 T. JEFFERSON L. REV. 103 (2013); Raymond Richard Neutra et al., *The Use and Misuse of Bradford Hill in U.S. Tort Law*, 8 JURIMETRICS J. 127 (2018).

to assert specific causation, and therefore the Bradford Hill indicia are useful in that context as well. In the legal context, and in specific cases, some of the indicia will be more important than others. It is important to emphasize that these factors are meant to be used as indicia or touchstones of causation. They can only assist a court's (or a jury's) judgment about general causation, not provide a definitive answer to that important question. These indicia are strength, consistency, specificity, temporality, biological gradient, plausibility, coherence, experiment, and analogy.[105]

*1. Strength of the Relationship*

A body of epidemiologic evidence showing a strong relationship between a purported cause and its effect suggests causation because a strong correlation makes it less likely that plausible confounders explain a relationship.[106] This is especially the case if several studies on different populations demonstrate a strong relationship. If the size of an effect is great enough, one may even argue that it is greater than the combined effect of plausible other causes.

When the strength of a relationship is modest, caution must be taken in interpreting the results. This is especially true in those case-control studies where the cases and controls are not drawn from the same defined population at risk for the outcome under investigation and, therefore, where internal validity threats such as selection bias and information bias may cast doubt on a causal interpretation.[107] At the other extreme, when a substantial body of epidemiologic evidence fails to find a substantial relationship, other evidence is unlikely to lead to the conclusion that there is a causal relationship. We elaborate on this point in the consistency discussion below.

A different and more complex strength question is posed when there is no epidemiologic evidence. Without epidemiologic or other human evidence, external-validity threats become much more significant due to the uncertainty of extrapolations from doses used *in vivo* and *in vitro* studies to human dose equivalents. To this we must add the external-validity threats arising from cross-species comparisons and, in the case of *in vitro* research, the lack of any metabolism effects. All of this points to the underlying fact that the strength of a relationship is of limited use in resolving external-validity threats due to possible effect modification by variables that differ between the populations or circumstances that have been studied and the ones targeted for inference. It is important to note that because many animal studies are true experiments, even more modest size effects

105. These indicia are not always mutually exclusive and may overlap in some circumstances.

106. Jeremy Howick et al., *The Evolution of Evidence Hierarchies: What Can Bradford Hill's 'Guidelines For Causation' Contribute?*, 102(5) J. ROYAL SOC. MED. 186, 187 (2009). This presumes that we have attempted to control for plausible confounders. When we fail to do so, even strong correlations may be misleading. *See* Kenneth J. Rothman & Sander Greenland, *Causation and Causal Inference in Epidemiology*, 95 AM. J. PUB. HEALTH 144 (2005).

107. Noel Weiss, *General Concepts in Epidemiology*, in FAIGMAN ET AL., *supra* note 11, § 23.37.

are likely to be causal, and even with respect to nonexperimental research, the greater the strength of a relationship the more likely the observed effect is causal.

*2. Consistency*

Consistency problems arise when studies of a similar type present mixed results. This may occur, for example, in epidemiologic research when some studies find substantial and statistically significant correlations between a substance and an injury while others do not, perhaps because they involved different populations.[108] Replication of early findings, ideally with improved methods, is a step most epidemiologists require before they would be comfortable with a causal interpretation of such a relationship.[109] However, in defining consistency, one must be careful to distinguish between differences in the actual effects suggested by data and differences in statistical significance of the observed effects in different studies, which may arise solely due to differences in sample sizes or measurement precision. To replace comparisons of observed relationships in different studies with comparisons of their *p*-values is a serious inferential error.

Consistency is also an indicium of causation when studies in one discipline, e.g., epidemiology, are consistent with studies in another area, e.g., *in vivo* studies. When a substance produces similar injuries in humans, as measured by epidemiologic research, and *in vivo* studies, this lends support to a causal interpretation. Contrariwise, lack of consistency or a failure to find a relationship across areas of study may be persuasive that there is no causal relationship—though lack of consistency must be interpreted with some caution.

Lack of consistency across domains of investigation requires a somewhat more extensive discussion when the apparent inconsistency involves epidemiologic research. Epidemiologic evidence may disagree with other evidence in two ways. First, it may be difficult to find an animal model that replicates the harm a substance does to humans. Recall the first animal model to be exposed to thalidomide did not suffer the limb-reduction defects typical of human fetus exposure.[110] Thus, the existence of consistency across disciplines supports a causal interpretation, but a lack of consistency does not necessarily provide strong evidence against a causal interpretation with respect to living humans.

Second, there may be evidence of harm to cells in an *in vitro* study or to some animal-toxicology models but a body of epidemiological work that fails to

---

108. *See, e.g.*, Pritchard v. Dow Agro Scis., 705 F. Supp. 2d 471, 488–89 (W.D. Pa. 2010), *aff'd*, 430 Fed. App'x 102 (3d Cir. 2011).

109. There are other ways to increase our confidence that there is a causal relationship. If a group of studies suggest a relationship, but the results in some studies are not statistically significant, meta-analyses may confirm a relationship that may have been too small to be detected in individual studies. *See* MICHAEL BORENSTEIN, INTRODUCTION TO META-ANALYSIS (2009).

110. *See supra* p. 25. Similarly, there apparently has been only partial success in finding animal models that mimic human response to inhaled tobacco smoke. Stephen S. Hecht, *Carcinogenicity Studies of Inhaled Cigarette Smoke in Laboratory Animals: Old and New*, 26(9) CARCINOGENESIS 1488 (2005).

reveal an injury to humans, as is the case with saccharin exposure.[111] Where epidemiological studies consistently fail to find a relationship between a relevant exposure and disease and an expert attempts to rely on other research to prove causation, most courts explicitly or implicitly adopt a hierarchy-of-evidence approach with epidemiologic (and clinical trial) evidence at the top of the hierarchy. Plaintiffs cannot proceed by ignoring these findings. This was the case in both the Bendectin litigation and the connective-tissue-disease, silicone-breast-implant litigation.[112] Similar positions have been adopted with respect to other exposures as well.[113]

*3. Specificity*

This indicium is thought by some to be the least diagnostic.[114] This is true if one understands specificity to mean that a causal relationship is stronger when a single putative cause produces a single specific effect. As many have noted, at least when we are looking for the effects of causes, the absence of this type of specificity does not undermine a causal conclusion. Cigarette smoke causes many diseases, and many of those diseases themselves have other causes.[115]

There are other ways to understand this indicium that are more helpful. The "second law" of toxicology concerns the specificity of toxic effects of individual substances. Harmful substances generally cause a limited range of specific injuries. For example, Vioxx, and to a lesser extent all NSAIDs[116] appear to have the potential to cause heart attacks and strokes but apparently not other injuries such as cancer.[117] As mechanistic toxicology acquires better understanding of the biological processes that produce an injury, and as toxicogenomics uncovers other biomarkers of exposure, specificity promises to take on additional importance as an indicium of causation.[118] Specificity also plays a role in animal studies. Animals exposed to a

---

111. NAT'L TOXICOLOGY PROGRAM, U.S. DEP'T OF HEALTH & HUMAN SERVICES, REPORT ON CARCINOGENS (PB2000107509) (9th ed. 2000). This and other examples of differences in metabolic processing of a chemical in the animal species and in humans demonstrate that positive carcinogenicity bioassays in rodents certainly suggest further research is needed. But one must guard against hasty extrapolations to humans before there is evidence that the animal in question and humans share similar biological mechanisms. Of course, the case for a human effect is much stronger if there are significant effects in several animal species and strains of laboratory animals.

112. *See* Richardson by Richardson v. Richardson-Merrell, Inc., 857 F.2d 823, 832 (D.C. Cir. 1988); Norris v. Baxter Healthcare Corp., 397 F.3d 878, 882 (10th Cir. 2005).

113. *See* Doe v. Ortho-Clinical Diagnostics, Inc., 440 F. Supp. 2d 465, 474 (M.D.N.C. 2006); Parker v. Mobil Oil Corp., 857 N.E.2d 1114 (N.Y. 2006); Blackwell v. Wyeth, 971 A.2d 235, 264 (Md. 2009); *In re* Accutane Litig., 191 A.3d 560, 592 (N.J. 2018).

114. Neutra et al., *supra* note 104.

115. Cigarette smoke is a complicated example because it is comprised of many chemicals with separate effects. But other substances for which this is not the case—e.g., alcohol and asbestos—also produce a variety of injuries.

116. NSAIDs are nonsteroidal anti-inflammatory drugs.

117. Michele Bally et al., *Risk of Acute Myocardial Infarction with Real-World NSAIDs Depends on Dose and Timing of Exposure*, 27(1) PHARMACOEPIDEMIOLOGY & DRUG SAFETY 69 (2018).

118. *See* Noel Weiss, *Can the Specificity of an Association be Rehabilitated as a Basis for Supporting a Causal Hypothesis?*, 13 EPIDEMIOLOGY 6 (2002).

toxic substance may suffer injuries different from those suffered by humans.[119] The discovery of a specific mechanism of injury may explain species differences.

Although Bradford Hill meant for specificity to refer to the injury caused by a substance, in the legal context it may also be useful to think of specificity more broadly as pertaining to the relationship between the individual's circumstance and the research cited in support of a causal relationship. This raises considerations of external validity. The subjects in a study may have been exposed to a substance somewhat different from the substance to which the plaintiff was exposed.[120] Specificity may also relate to dosage. The level of exposure to which an individual is exposed may differ from the exposure level in relevant research.[121]

Specificity is inherently a problem with *in vivo* and *in vitro* research because of external-validity problems. Nevertheless, many courts have stated that a person does not need epidemiologic evidence to show general causation.[122] This situation arises where experimentation is precluded due to ethical or other reasons and when it is also difficult to conduct epidemiologic research with sufficient statistical power due to the rarity of the disease in question.[123] However, large samples are not always necessary for successful and important epidemiologic research.[124]

The absence of epidemiological evidence, however, does not mean that any toxicologic evidence will suffice. Animal studies usually pose fewer external-validity problems than some other types of toxicological data, and courts are more likely to reject evidence that is based solely on *in vitro* studies or on a structure-activity analysis.[125] However, courts are rarely confronted with pure *in vitro* cases

---

119. General Electric v. Joiner, 522 U.S. 136, 144 (1997).

120. *General Electric v. Joiner* provides an example. Some subjects in one of the epidemiology studies referred to in the opinion had been exposed to mineral oil, and the study did not mention PCBs. *Id.* at 145–46.

121. When courts do exclude an expert's testimony that is based in part on epidemiological research, it is almost always the case that the study or studies in question fail on one or more of the aforementioned factors. *See, e.g.*, Magistrini v. One Hour Martinizing Dry Cleaning, 180 F. Supp. 2d 584 (D.N.J. 2002), *aff'd*, 68 Fed. App'x 356 (3d Cir. 2003) (exposure to different chemical); Est. of Mitchell v. Gencorp, Inc., 968 F. Supp. 592, 600 (D. Kan. 1997), *aff'd*, 165 F.3d 778 (10th Cir. 1999) (outcome was a different disease).

122. This position was affirmed by three different circuits in cases involving Parlodel. *See* Glastetter v. Novartis Pharms. Corp., 252 F.3d 986 (8th Cir. 2001); Hollander v. Sandoz Pharms. Corp., 289 F.3d 1193 (10th Cir. 2002); Rider v. Sandoz Pharms. Corp., 295 F.3d 1194 (11th Cir. 2002); Dunn v. Sandoz Pharms. Corp., 275 F. Supp. 2d 672 (M.D.N.C. 2003); *see also In re* Meridia Prods. Liab. Litig., 328 F. Supp. 2d 791, 800 (N.D. Ohio 2004), *aff'd*, 447 F.3d 861 (6th Cir. 2006); *In re* Heparin Prods. Liab. Litig., 803 F. Supp. 2d 712 (N.D. Ohio 2011).

123. *See* McClellan v. I-Flow Corp., 710 F. Supp. 2d 1092, 1114, (D. Or. 2010).

124. Arthur L. Herbst et al., *Adenocarcinoma of the Vagina: Association of Maternal Stilbestrol Therapy with Tumour Appearance in Young Women*, 284(16) NEW ENG. J. MED. 878 (1971); Peter B. Imrey et al., *Outbreak of Serogroup C Meningococcal Disease Associated with Campus Bar Patronage*, 143(6) AM. J. EPIDEMIOLOGY 624 (1996).

125. *See* McClain v. Metabolife Int'l, Inc., 401 F.3d 1233, 1245 (11th Cir. 2005); Siharath v. Sandoz Pharms. Corp., 131 F. Supp. 2d 1347, 1364 (N.D. Ga. 2001), *aff'd*, 295 F.3d 1194 (11th Cir. 2002).

because it is unlikely that a toxicologist would conclude that a substance caused harm to individuals based solely on this type of research.

*4. Temporality*

Temporality is a universally accepted indicium. Indeed, the existence of a temporal order—the effect must follow the putative cause—is generally thought to be an essential criterion before one can attribute causation.[126] Three caveats are in order. First, some cross-sectional research designs make it difficult to assess temporal order.[127] Second, the fact that a putative cause precedes an effect is generally not, by itself, persuasive evidence of a causal relationship. In the next Part we shall see that many courts have rejected expert testimony that seems to rely exclusively on the fact that the alleged cause preceded the effect. Third, even though an effect may follow an alleged cause, it may follow too soon or too late for the alleged cause to be plausible as the cause, based on existing research. No one contracts mesothelioma one year after exposure to asbestos.[128] One year after that exposure would be too soon to become sick from that disease, and so asbestos exposure one year prior to onset of mesothelioma would not plausibly be causal for that illness.[129]

*5. Biological Gradient*

The first "law of toxicology," that the dose makes the poison, implies among other things that the incidence and/or the severity of the disease increases as exposure rises. This indicium is met when a greater exposure leads to a more serious manifestation of a disease. This is the case, for example, with the level of exposure to asbestos and the severity of an individual's asbestosis. The criteria are also met when a substance causes a disease, such as mesothelioma, that does not vary in its severity but that does occur with greater frequency as the level of exposure increases.[130] The apparent complete absence of a biological gradient in either of these senses generally argues against causation, with the caveat that a sufficiently wide range of doses must have been studied.[131]

As we noted above, with respect to many exposures and many types of injury, it is commonly thought that there is some threshold dose below which the substance does not cause harm. Recall that toxicology research typically attempts to

---

126. For a case where temporal order is unclear, see *Guinn v. AstraZeneca Pharmaceuticals LP*, 602 F.3d 1245, 1254 (11th Cir. 2010).

127. For example, if in a study conducted at a single point in time, we observe that individuals taking a certain antipsychotic drug have a higher suicide rate than individuals not taking the drug, we cannot know if the drug caused suicide or whether suicidal tendencies are the reason these individuals began taking the drug. To sort out this question, we need studies that observe people over time.

128. FAIGMAN ET AL., *supra* note 11, § 22.24.

129. The second and third caveats are particularly relevant when we turn to the problem of specific causation. We discuss them further there.

130. *See* FAIGMAN ET AL., *supra* note 11, § 22.24 (providing an example of a dose-response relationship figure).

131. There may be situations, however, where the range of exposure in the research is too narrow to detect a dose–response relationship that would emerge if there were a greater range of exposures.

estimate the dosage below which there is no observable effect. Similarly, epidemiological research may indicate adverse effects for higher-dose exposure but no significant effect for those exposed to lower doses. This can reflect a true threshold effect or just insufficient sample sizes to detect low-dose effects but regardless causes expert-testimony admissibility problems for plaintiffs exposed to lower doses.[132] However as we noted above, with respect to some substance/injury relationships, it is thought that there is no safe threshold.[133] Cancer is the injury for which it is most frequently thought that there is no safe threshold, but even here the mechanism of injury may lead to a different conclusion.[134]

*6. Plausibility*

A useful way of thinking about biological plausibility is to ask whether a hypothesis that a certain exposure causes a certain disease makes sense given what we know about other exposures and other injuries and about the biological mechanisms through which an exposure causes an injury. Plausibility is perhaps most useful as an exclusionary criterion. For example, a teratogen cannot cause certain types of fetal injuries if there is persuasive evidence that it cannot cross the placental barrier.

Plausibility has several dimensions. Plausibility is often a question when discussing *in vivo* research. A human injury is more plausible if we observe injuries in animals experiencing a similar exposure.[135] Plausibility is also an issue when one compares exposure to one substance with exposures to other substances with a similar chemical structure. If the other substances are known to cause injury, this lends some credence to the possibility that the substance under consideration also may cause such injuries. Often, however, this so-called "structure activity" evidence is of limited probative value. Recall the earlier discussion in footnote 70 of the different carcinogenetic effects of benzene, toluene, and xylene.[136] Understanding why these differences exist provides an example of perhaps the most frequent use of plausibility evidence today: proposing the biological mechanism by which an exposure causes injury.

---

132. *In re* Bextra & Celebrex Mktg. Sales Pracs. & Prod. Liab. Litig., 524 F. Supp. 2d 1166, 1180 (N.D. Cal. 2007) (distinguishing between a 200 mg/d dose of Celebrex and a 400 mg/d dose; excluding expert testimony that the lower daily dose could cause heart attacks and strokes but permitting expert to testify that the higher daily dosage could cause these injuries); *In re* Lipitor (Atorvastatin Calcium) Mktg., Sales Pracs. & Prods. Liab. Litig., 174 F. Supp. 3d 911 (D.S.C. 2016) (permitting plaintiff's epidemiologists to testify that 80 mg/d of Lipitor could cause Type 2 diabetes but excluding as unreliable their testimony that 10 mg/d could do so).

133. For a discussion of extrapolations when we believe there is no threshold dose, see FAIGMAN ET AL., *supra* note 11, § 22.29.

134. *See* Chlorine Chemistry Council v. Env't Prot. Agency, 206 F.3d 1286 (D.C. Cir. 2000) (rejecting the EPA's assumption that there is no safe threshold with respect to carcinogenic effects of chloroform in drinking water).

135. Here, the nature of the exposure and the injury play a role as was the case in the mouse study discussed in the Supreme Court's *Joiner* opinion. Gen. Elec. Co. v. Joiner, 522 U.S. 136, 144 (1997).

136. *See supra* note 70.

Plaintiffs frequently offer evidence of this type of biological plausibility to support their causal argument. For example, in *In re Neurontin Marketing, Sales Practices, and Products Liability Litigation*,[137] the plaintiffs alleged that they, or their decedents, suffered suicide-related injuries when their physicians prescribed the anti-epileptic drug Neurontin. An important part of their argument was that the drug reduced serotonergic levels, a known correlate with mood changes and depression. The court agreed that this analysis strengthened the plaintiffs' case.[138]

It is worth noting, however, that the plaintiffs had other, more direct epidemiologic evidence of a correlation between antiepileptic drugs and suicide. When other, more direct evidence is unavailable, biologic-plausibility evidence has been found to be less persuasive.[139]

*7. Coherence*

Coherence refers to whether the causal hypothesis is supported by or is in conflict with other facts about the history and distribution of the illness. The coherence indicium often overlaps with the plausibility indicium. In cases, the terms often seem to be used interchangeably to discuss things such as whether an epidemiologic research finding is supported by similar results in animal studies. As is the case with biologic plausibility, courts are most receptive to the use of this guideline when there is other, more direct evidence of causation.[140]

*8. Experiment*

Bradford Hill's comments about the experiment indicium seem framed in terms of one particular quasi-experimental design, sometimes referred to as a "challenge/dechallenge design." Our belief that a substance caused an injury is strengthened if the injury ceases when the cause is removed. Allergists frequently use this design to test for what allergen is causing the patient's symptoms.[141] Of course, this technique is useful only for acute responses that are reversible upon withdrawal of a suspect cause.

For the purposes of this Article, however, it is useful to consider the whole range of experimental designs—designs in which the exposure is systematically manipulated by the researcher. Properly conducted, true experiments come closer than other approaches to avoiding challenges to internal validity and supporting causal explanations. While quasi-experiments are not as good in resolving internal-validity issues, they are largely better than designs without any manipulation of the

---

137. 612 F. Supp. 2d 116, 121 (D. Mass. 2009).

138. *Id.* at 158. The plaintiffs in this case also argued for biological plausibility based on *in vitro* and *in vivo* research and because the active ingredient had a chemical structure similar to other antiepileptic drugs. *Id.* at 145–48.

139. In several cases, courts have stated that plausibility arguments, standing alone, are insufficient to support a causation conclusion in the absence of evidence showing a relevant association between an exposure and an injury. Some of these cases are discussed in *In re Lipitor (Atorvastatin Calcium) Marketing, Sales Practices, and Products Liability Litigation*, 174 F. Supp. 3d 911 (D.S.C. 2016). Such a conclusion turns, however, on the quality of mechanism evidence supporting an expert's conclusion.

140. *See id.*

141. Jinoos Yazdany et al., *Allergy Testing*, *in* CURRENT MEDICAL DIAGNOSIS AND TREATMENT 20–55 (Maxine A. Papadakis et al. eds., 2021).

alleged cause.[142] But causal inferences from even the best experiments may be open to external-validity threats.

*9. Analogy*

Are there known causal relationships of similar exposures to similar illnesses? The probative value of this indicium turns on both similarity of the substances and the similarity of the injury.[143] The idea of analogy may fit many other situations as well. All use of animal studies to show human causation could be thought of as an argument by analogy.

In summary, and to repeat our cautionary comments at the outset of this Part, these nine indicia are very far from litmus tests of general causation. Even the strongest correlation can be spurious, and the importance of factors such as consistency may turn on the exact nature of an inconsistency. However, the presence or absence of indicia do properly influence our willingness to believe a relationship is causal. They do so because they alter the vulnerability of a body of data to alternative explanations from internal and external validity threats. As this discussion also demonstrates, the indicia are not of equal importance in helping one make a causal attribution. Temporal order is essential. Experiment, strength, and consistency are also often key. Plausibility will play an increasingly important role with a firmer understanding of the precise biological mechanisms by which a substance causes an illness. In many situations, the absence of a biological gradient can be an important indicium arguing against a causal conclusion. As we discuss below, these and the other Bradford Hill indicia may play quite similar roles in specific causation inquiries.

## IV. RULING OUT: SPECIFIC CAUSATION

The third step in the three-step process of differential etiology is using "standard diagnostic techniques by which doctors normally rule out alternative causes" to reach a conclusion as to which cause is most likely.[144] General causation addresses the issue of whether a substance *can* cause the plaintiff's condition; specific causation addresses the issue of whether the substance in question *did* cause the plaintiff's condition. We agree with the reference manual's statement that the logic of Step Three is sound: "[E]liminating other known and competing causes increases the probability that a given individual's disease was caused by exposure to the agent."[145] However, therein lies the problem. While courts set forth the logic of differential etiology, they offer no real guidance about the details of Step Three.[146]

---

142. On the general topic of quasi-experimental designs, see DONALD T. CAMPBELL & JULIAN C. STANLEY, EXPERIMENTAL & QUASI-EXPERIMENTAL DESIGNS FOR RESEARCH (1963); THOMAS D. COOK & DONALD T. CAMPBELL, QUASI-EXPERIMENTATION: DESIGN & ANALYSIS ISSUES FOR FIELD TESTING (1979).

143. Soldo v. Sandoz Pharms. Corp., 244 F. Supp. 2d 434, 515 (W.D. Pa. 2003).

144. Best v. Lowe's Home Ctrs., Inc., 563 F.3d 171, 179 (6th Cir. 2009).

145. Green et al., *supra* note 101, at 617.

146. By our count, over 70 cases have repeated the phrase "standard diagnostic techniques," but they do not provide any overall guidance as to the method one should use in ruling out alternative causes.

This is not to say that the courts have no concept of a good or a bad differential-etiology analysis. Many opinions provide coherent discussions as to whether an expert has good grounds for a differential etiology. What is lacking is a systematic approach for assessing the merits of a differential etiology.

Perhaps it goes without saying that in a post-*Daubert* world an expert must present some justification for her opinion. Under the older *Frye* rule,[147] some courts explicitly permit an expert to testify as to specific causation based solely on her opinion or "clinical experience."[148] One import of the *Daubert* line of cases is that this type of *ipse dixit* will not do. What then constitutes an acceptable analysis?

Recall that Rule 702 of the Federal Rules of Evidence[149] now requires that expert testimony must be based on sufficient facts and data, and that the testimony be the product of reliable principles and methods. In the specific-causation context, the rule thus imposes two requirements: sufficient data with which to reach a conclusion that the substance under question is the cause of the plaintiff's injury and, equally important, a reliable method to address the effects of alternative possible causes.

This Part uses the Bradford Hill indicia and an analysis of validity to flesh out these two requirements and to provide a systematic approach to those factors that courts should consider in assessing specific causation. In short, what Bradford Hill did for general causation in epidemiology we hope to do for specific causation. As is the case in our earlier general-causation discussion, these indicia are meant as guidelines. Strength in certain dimensions can make up for weaknesses in others, but there is no algorithm that points to a certain conclusion. Throughout this Part we will make occasional references to the admissibility of specific-causation testimony.

It is worthwhile to emphasize several key caveats at the start. First, if both of the first steps of differential etiology are not satisfied—correct diagnosis and general causation—there can be no specific-causation determination. The great majority of courts agree on this point.[150] Second, as discussed in greater detail below, what is known about general causation—both as to the cause to be ruled-in as well as to the alternative cause(s) to be ruled-out—has a direct impact on what might be said about specific causation. Third, as noted, these specific-causation guidelines are just that: guides to informed decision-making on what is almost invariably a task fraught with uncertainty. Finally, although scientific reasoning can help reduce this uncertainty, the question of where to set the uncertainty threshold for admission of evidence as proof is a question of policy as well as science. We consider this point in Part V.

---

147. The *Frye* rule is named after the opinion in *Frye v. U.S.*, 293 F. 1013 (D.C. Cir. 1923), which established the "general acceptance" test for admissibility.

148. *See, e.g.*, Kuhn v. Sandoz Pharms. Corp., 14 P.3d 1170, 1179 (Kan. 2000); Marsh v. Valyou, 917 So. 2d 313, 327 (Fla. Dist. Ct. App. 2005), *review granted*, 940 So. 2d 1125 (Fla. 2006), *decision quashed*, 977 So. 2d 543 (Fla. 2007).

149. *See supra* note 9 (providing the full text of Rule 702).

150. *See* FAIGMAN ET AL., *supra* note 11, § 21.6.

### *A. Specificity of Cause and Condition*

The specific-causation inquiry requires us to focus on the causes of effects. This analysis poses special problems. A typical tort example may help illustrate the nature of the difficulty.

Suppose Ann took the drug benfluorex as an appetite suppressant. At some later time, she develops valvular heart disease. In her case, and in each case where an individual is suing the source of the injury, we know that she has been exposed to the proposed cause, and we know she has suffered the relevant effect. Did the exposure cause the disease? When there are no other putative causes under consideration, the "cause of an effect" question may be stated as follows: "If Ann had not taken benfluorex, and other things being equal, would she still have developed valvular heart disease?"[151] This can also be phrased as addressing, in terms of formal logic, the question of causal necessity and, equivalently in legal terms, the "but for" assertion: "But for the benfluorex, Ann would not have developed valvular heart disease." The answer requires a form of counterfactual reasoning because we cannot observe the alternative, i.e., the situation where Ann did not take benfluorex.

Resolving the specific causation question resolves down to a set of questions. First, even when there is only one "cause" under consideration, e.g., benfluorex, what is the evidence supporting the argument that Ann's outcome is an instance of the alleged causal relationship rather than an unrelated natural ("idiopathic") development? Second, in the common scenario where there are competing rival causes, what is the evidence supporting the argument that Ann's injury is an instance of some other causal relationship, such as a family history of valvular heart disease? Intuitively, causal determination turns on the comparative likelihood that her injury is a consequence of the alleged cause versus other potential causes. However, in the presence of multiple possible causes, it is not clear how to assess these likelihoods: in particular, using the "but for" approach for each cause may well implicate several of the possible causes simultaneously. In consequence, when "probabilities of causation" for the different suspected causes are computed based on this approach, they may well sum to more than one.

Both as a matter of common sense and scientific research, there are circumstances in which there is an overwhelming basis for connecting a cause to a specific condition. When someone with two healthy legs is involved in a car accident and emerges with a broken leg, we can be fairly confident that the accident was the cause.[152] Simply put, in the calculation of ruling-in one cause and ruling-out others, if there are no plausible causes to rule-out, specific causation is easier.

The cases where there are most likely to be no viable alternative causes are those involving an immediate effect following an exposure. In *Heller v. Shaw*

---

151. The example is taken from Dawid & Musio, *What Can Group Level Data Tell Us About Individual Causality?*, *in* STATISTICS IN THE PUBLIC INTEREST—IN MEMORY OF STEPHEN E. FIENBERG (A. Carriquiry et al. eds., Springer 2021).

152. *See* Bonner v. ISP Techs., Inc., 259 F.3d 924 (8th Cir. 2001); Marcum v. Adventist Health Sys./W., 193 P.3d 1, 2 (Or. 2008) (en banc); McClellan v. I-Flow Corp., 710 F. Supp. 2d 1092, 1094–95 (D. Or. 2010).

*Industries, Inc.*, Judge Becker focused on this aspect of the temporal relationship dimension of toxic tort cases.[153] When an acute exposure is immediately followed by an injury with no other plausible causes, courts are much more likely to permit an expert to testify as to specific causation based in large part on temporal order. Specific causation most often becomes problematic when there is a significant passage of time between exposure and the emergence of an injury.

Even when the injury does not immediately follow exposure, there are cases where a specific-causation determination is straightforward due to the absence of other plausible causes. One well-known group of such cases are those that involve so-called signature diseases. The two best known signature diseases in the toxic tort arena are two ailments that arise from exposure to asbestos: asbestosis and mesothelioma. The former only occurs due to asbestos exposure, and while there are mesothelioma cases where there has been no known asbestos exposure, such cases are very rare.[154] When an individual presents with one of these diseases, it is almost certainly due to asbestos exposure.[155] However, outside the asbestos arena, signature diseases are quite rare, and the signature-disease approach to specific causation is virtually never applied to other ailments.[156]

For some illnesses that are not signature diseases, determining etiology may still be relatively straightforward. In some cases, researchers can use experimental or quasi-experimental methods to exclude other causes in the individual case. An allergist, for instance, can experiment with different allergens to identify the offending substance or substances.[157] In other cases, confidence in specificity is gained when all other plausible causes are convincingly ruled out.[158]

The cases discussed above are easier to resolve: if there is evidence of general causation, the injury is likely a consequence of the alleged cause because instances in which the same injury is associated with other causes are so much rarer. In more complex circumstances, cases become more difficult.

---

153. 167 F.3d 146, 154 (3d Cir. 1999).

154. Betz v. Pneumo Abex, LLC, 44 A.3d 27, 51 (Pa. 2012)

155. The asbestos cases do pose difficult causal questions because there are often multiple potential defendants whose product may have injured the plaintiff. In these situations, there is no question about the cause, but there are very difficult questions of whether a particular defendant's product should constitute a legal cause of the injury. *See* Joseph Sanders, *The "Every Exposure" Cases & the Beginning of the Asbestos Endgame*, 88 TUL. L. REV. 1153 (2014).

156. They do include silicosis (caused by breathing tiny bits of silica) and manganism, a disease caused by exposure to manganese. *See* Betsy J. Grey & Gary E. Marchant, *Biomarkers, Concussions, and the Duty of Care*, 2015 MICH. ST. L. REV. 1911, 1926 (2015). Note that research suggests that traumatic encephalopathy (CTE) is a signature disease caused by repetitive head injuries. Vaginal adenocarcinoma in young women appears to be a signature disease associated with maternal use of DES.

157. Yazdany et al., *supra* note 141, at 20–55.

158. *See* Yarchak v. Trek Bicycle Corp., 208 F. Supp. 2d 470, 498 (D.N.J. 2002); *In re* Fosamax Prods. Liab. Litig., 688 F. Supp. 2d 259 (S.D.N.Y. 2010); Tedder v. Am. Railcar Indus., Inc., 739 F.3d 1104, 1108–09 (8th Cir. 2014).

### *B. The Strength of the Relationship Between a "Cause" and an "Effect": Forms of Empirical Proof*

In this and the following Section we discuss situations where inferences about specific causation necessarily depend on the foundation of empirical work that supports general causation, including clinical trials, epidemiological research, various areas of toxicology, and genetic information. Each type of evidence must, however, be assessed for internal and external validity.[159]

Assuming adequate validity, one important question is when, if ever, general-causation results might directly imply specific causation. The most frequently discussed example is presented when there is a body of on-point epidemiology demonstrating a relative risk greater than 2.0, implying that if this relative risk accurately reflects causal mechanisms, then the fraction of new cases among exposed persons actually resulting from this exposure (or in epidemiological jargon, the true "attributable fraction among the exposed" or "individual attributable fraction") exceeds one half.[160] Scholars have long debated whether proof of a relative risk greater than 2.0 is sufficient to prove specific causation by a preponderance of the evidence.[161] Courts have been receptive to this assertion,[162] and we agree that theoretically this intuition has appeal.[163]

But before jumping to the somewhat simplistic, albeit intuitive, view that a doubling of the risk supports a more-probable-than-not determination, the twin pillars of validity—internal and external—must be evaluated. As regards internal validity, we must emphasize that any statistical statement is only as good as the research methods used to obtain it. A relative risk greater than 2.0 is less persuasive if there is only a single study supporting it; if the methods used in multiple studies suffer from significant validity problems; if other bodies of research, such as toxicological studies, fail to support the underlying plausibility of the cause-and-

---

159. Suppose researchers publish a study that purports to show that illness X is uniquely associated with substance Y. A close read of the article, however, indicates that the researchers studied only male subjects (of whom there were only a small number) and failed to employ double-blind methods, limited the number of illnesses evaluated, employed less-than-ideal-statistical methods, and no other researchers had yet replicated their results. The claim of individuality would be severely undermined by the validity threats associated with the general research program.

160. By "on-point," we mean to exclude studies that present substantial internal- and/or external-validity threats.

161. *See* Dow Pharms., Inc. v. Havner, 953 S.W.2d 706, 716 (Tex. 1997). *See generally* Susan Haack, *Risky Business: Statistical Proofs of Specific Causation*, *in* EVIDENCE MATTERS: SCIENCE, PROOF, & TRUTH IN THE LAW (2014); Carruth & Goldstein, *supra* note 91, at 203–04.

162. *See, e.g.*, Magistrini v. One Hour Martinizing Dry Cleaning, 180 F. Supp. 2d 584, 591 (D.N.J. 2002); Daubert v. Merrell Dow Pharms., Inc., 43 F.3d 1311, 1321 (9th Cir. 1995); Merrell Dow Pharms., Inc. v. Havner, 953 S.W.2d 706 (Tex. 1997); *In re* Silicone Gel Breast Implants Prods. Liab. Litig., 318 F. Supp. 2d 879, 893 (C.D. Cal. 2004); Est. of George v. Vt. League of Cities & Towns, 993 A.2d 367, 375 (2010); Johnson & Johnson Talcum Powder Cases, 249 Cal. Rptr. 3d 642 (Cal. Ct. App. 2019).

163. *See* Dawid et al., *supra* note 8, at 374–78; A. Philip Dawid et al., *From Statistical Evidence to Evidence of Causality*, 11(3) BAYESIAN ANALYSIS 725 (2016).

effect claim; or if the lower confidence bound for the true relative risk from a comprehensive meta-analysis falls well below 2.0.

As regards external validity, a pivotal issue in making specific causation judgments is whether the extant research can be applied to the individual in question. An obvious instance of this issue is whether a demonstrated effect in mouse studies can be applied to humans; a less obvious instance is whether a doubling of risk in studies of human adults can be applied to children.

A more difficult situation arises when epidemiologic data strongly suggest general causation but with relative risk below 2.0, so that the observed individual attributable fraction is below one half. Judge Kozinski argued in his *Daubert* remand opinion that a relative risk below 2.0 suggests it is less likely than not that the plaintiff's injury was caused by the substance in question.[164] Some courts have made a relative risk greater than 2.0 a litmus test for admissibility. The most extreme version of this approach may be found in the Texas Vioxx case, *Merck & Co. v. Garza*,[165] but *Garza* is not alone.[166] Only a few courts have adopted this bright-line rule, and this unqualified position is open to question. While statistical analyses provide support for a court's inclination to regard a relative risk above 2.0 as meeting the civil preponderance-of-the-evidence burden of persuasion, the opposite is not the case. A finding that a relative risk falls short of 2.0 does not necessarily imply that the probability of causation is less than one half. This is because relative risk does not determine probability of causation but only a lower bound for it.[167] Moreover, statistical statements such as relative risk and *p*-values constitute an important component but only one part of the research process: they cannot be used alone as talismans of admissibility, and their relevance to the admissibility decision is not straightforward.[168] For example, one can legitimately question whether an odds ratio of 1.9 or 2.1 is as meaningful when found only in case-control studies. Compare this to a 2.1 relative risk in cohort studies, which are generally less susceptible to internal validity threats, and thus the cohort study results are usually regarded, in principle and on balance, as better founded.[169] However, the extent to

---

164. 43 F.3d at 1321.

165. 347 S.W.3d 256, 265–66 (Tex. 2011).

166. *See* Sanderson v. Int'l Flavors & Fragrances, Inc., 950 F. Supp. 981, 1000 (C.D. Cal. 1996); Hall v. Baxter Healthcare Corp., 947 F. Supp. 1387, 1403 (D. Or. 1996).

167. Dawid et al., *supra* note 8, at 377.

168. Ronald L. Wasserstein & Nicole A. Lazar, *The ASA Statement on p-Values: Context, Process, and Purpose*, 70(2) AM. STATISTICIAN 129 (2016).

169. For example, on the question of whether genital application of talc-based baby powder causes ovarian cancer, compare Katie M. O'Brien et al., *Association of Powder Use in the Genital Area with Risk of Ovarian Cancer*, 323 J. AM. MED. ASS'N 49 (2020) (a large-pooled cohort study), *with* Kathryn L. Terry et al., *Genital Powder Use and Risk of Ovarian Cancer: A Pooled Analysis of 8,525 Cases and 9,859 Controls*, 6(8) CANCER PREVENTION RSCH. 811 (2013) (a case-control study), *and* Ross Penninkilampi & Guy D. Eslick, *Perineal Talc Use & Ovarian Cancer: A Systematic Review and Meta-Analysis*, 29 EPIDEMIOLOGY 41 (2018) (primarily based upon case-control studies). The latter two studies may suffer from some recall bias. *See* Britton Trabert, *Body Powder & Ovarian Cancer Risk—What is The Role of Recall Bias?*, 25(10) CANCER EPIDEMIOLOGY BIOMARKERS & PREVENTION 1369 (2016).

which a research design achieves its potential is determined by how well the design is executed.

Given these caveats, how may courts address this situation? Several approaches are available, sometimes in combination. Each employ one or more Bradford Hill indicia. First, it may be possible to demonstrate that the plaintiff was exposed to a larger dose or a greater exposure than the subjects in the studies, and an expert might conclude that, had the subjects been exposed at this level, the relative risk would have been greater than the relative risk threshold of 2.0. Of course, the opposite may be true, in which case the best estimate of the strength of the relationship between exposure and injury in those exposed to the same level as the plaintiff would be lower than observed in the research. In this situation, even if epidemiologic research indicated a relative risk greater than 2.0, if the plaintiff's exposure were significantly less, then one may reasonably conclude that the relative risk for this individual is less than 2.0. If the exposure is small enough, one may even conclude there is negligible evidence of general causation.[170]

A second approach, recognizing the indeterminacy of any estimate of the strength of a relationship, allows some relaxation of the 2.0 relative-risk threshold. Courts may allow testimony based on epidemiological research that all agree supports a relative risk of 1.9.[171] This approach takes into account that the true population value of the relative risk is never exactly known, but only estimated within a range of accuracy that may generally be represented by a confidence interval, standard-error estimate, [172] or Bayesian credible interval. [173] This is especially true if one can point to aspects of the epidemiologic studies, such as a potential misclassification error that could have led to an attenuation of the relationship.[174] Of course, this approach is inappropriate if the relative risk is small,

---

170. *See Merck*, 347 S.W.3d at 266. Mr. Garza took 25mg a day of Vioxx for 25 days.

171. *In re* Hanford Nuclear Rsrv. Litig., 292 F.3d 1124, 1137 (9th Cir. 2002) ("We agree with the Third Circuit that the validity of a claim should not depend on whether a plaintiff was exposed to a fraction of a rem lower than the "doubling dose."); *see also* Grassis v. Johns-Manville Corp., 591 A.2d 671, 676 (N.J. Super. 1991).

172. FAIGMAN ET AL., *supra* note 11, § 23:39.

173. A Bayesian credible interval is an interval including a specified fraction, e.g., 95%, of the posterior probability distribution of an unknown parameter of a population or process, conditional on a presumed (prior) probability distribution of that parameter and on data observed from other probability distribution(s) indexed by that parameter. A 95% confidence interval, in contrast, is a one-sided or two-sided interval from a data sample with 95% probability of bounding a fixed, unknown parameter, for which no nondegenerate probability distribution is conceived, under specified assumptions about the data distribution. These types of intervals, neither of which is unique, play analogous roles in Bayesian and frequentist inference. They may under some conditions be numerically similar but have very different philosophical bases. Credible intervals may alternatively be chosen to have maximum posterior density, to center on the posterior mean, or to exclude equal posterior probability at lower and upper ends. *See* EMMANUEL LESAFFRE & ANDREW B. LAWSON, BAYESIAN BIOSTATISTICS 47–50 (2012).

174. On the problem of misclassification, see JOSEPH SANDERS, BENDECTIN ON TRIAL 80–82 (1998).

e.g., 1.1 or 1.2, even if one is persuaded that such a small relative risk is admissible evidence of general causation.[175]

Third, some courts have argued that a relative risk less than 2.0 may be supplemented with other, non-epidemiological evidence of causation.[176] This approach raises the question as to what sort of evidence could play this role? For example, does the existence of toxicological evidence indicating that a substance causes an injury in some mammals lend independent support to an expert's causation testimony in the face of epidemiological evidence indicating a relative risk less than 2.0? With respect to general causation, the answer is yes. The "third law" of toxicology is that humans are animals, meaning that the study of animals can provide useful insight into effects in humans.[177] Such information speaks to several Bradford Hill indicia, including consistency, plausibility, and coherence. With respect to specific causation, however, the answer is more nuanced and turns in part on the quality and relevance of the two types of studies. The argument for supplementing epidemiology with evidence on other species is strengthened if one can point to aspects of the epidemiologic research that could lead to an attenuation of the relationship. The expert's argument is also strengthened insofar as the substance and injury are similar across species,[178] an application of Bradford Hill's consistency indicium, and where there is mechanistic-toxicological evidence that describes a toxic pathway shared by humans, a form of plausibility evidence. In sum, the implications of this type of toxicological evidence for specific causation are best framed in terms of their contributions to satisfying and deepening the relevance of the Bradford Hill indicia for general causation to the specific case.[179]

There is no formula for when such toxicologic evidence can tip the scales on the question of specific causation. One can say, however, that this type of evidence is most helpful when epidemiologic research suffers from various types of internal-validity threats or when the mechanism of injury is not well understood. It is less helpful when the research on human subjects suffers from less-serious threats to internal validity. The problem is that even well-designed toxicologic research labors under external validity threats when applied to humans. There is no formula for extrapolating the strength of an association in an *in vivo* experiment conducted at relatively high dosage to the strength of a relationship between a substance and a human illness at substantially lower dosages. If there is a large, well-designed body

175. Note that this was the situation in the Bendectin cases. The best estimates of relative risk or odds ratio were 1.1. *Id.* For a more recent case making the same point, see *In re Viagra (Sildenafil Citrate) and Cialis (Tadalafil) Products Liability Litigation*, 424 F. Supp. 3d 781, 788 (N.D. Cal. 2020).

176. Landrigan v. Celotex Corp., 605 A.2d 1079, 1087 (1992); *In re Hanford*, 292 F.3d at 1135–36.

177. Goldstein & Henifin, *supra* note 77, at 401.

178. *See* Xanthi Pedeli et al., *Risk Assessment of Diesel Exhaust and Lung Cancer: Combining Human and Animal Studies After Adjustment for Biases in Epidemiological Studies*, 10 ENV'T HEALTH 30 (2011) (discussing a method to combine animal and human studies to produce a better estimate of risk); Douglas L. Weed, *Weight of Evidence: A Review of Concept and Methods*, 25 RISK ANALYSIS 1545 (2005).

179. Johnson & Johnson Talcum Powder Cases, 249 Cal. Rptr. 3d 642, 671–72 (Cal. Ct. App. 2019) (providing a useful example of how an expert may supplement epidemiologic studies having risk estimates less than 2.0 with other evidence).

of on-point epidemiology indicating a small relative risk, e.g., 1.02, it is not clear how any animal data could cause one to substantially alter the best estimate of a human effect to reach a more-likely-than-not threshold.

Some genetic information may also prove useful in this situation. For legal purposes, one must distinguish among markers of susceptibility, markers of an injury, and markers of exposure.[180]

Recall also our earlier discussion of *Bowen v. E.I. Du Pont De Nemours and Co., Inc.*,[181] where the plaintiff's claim was badly compromised when it was discovered she had a biomarker for an injury, CHARGE syndrome, different from the injury she was claiming. This was a marker for a disease, not an exposure. Biomarkers of disease can play an important part in cases where the diagnosis is in dispute as it was in the *Bowen* case.[182] But the *Bowen* diagnosis did more. It not only specified the disease, but it also indicated that a genetic defect was the overwhelmingly most-likely cause of the injury.

As in the *Bowen* case, the most useful biomarker would be one that would allow us to define signature diseases. That is, the marker would be able to differentiate injuries with known multiple causes into subsets within which everyone with the injury and the marker is known to have been exposed to the same putative cause. This does not guarantee there are no other potential causes of this effect, but it would almost certainly be admissible evidence on specific causation.[183]

### *C. The Presence of Competing Factors*

Thus far, we have focused on overly simplified situations where the question is, with respect to a given putative cause, what evidence is available to say that it is or is not more likely than not that the effect was produced by this cause? In point of fact, of course, even when an opinion does not discuss other potential causes, there are no cases (with the possible exception of signature diseases) where we can be certain there is a single cause. In this next part of this Article, we introduce other potential causes. These may come from many sources, including one's genetic makeup, lifestyle, preexisting health situation, or drug or other environmental exposures, to give but some examples. Thus, a careful differential etiology must often assess the strength of potential alternative causes of the plaintiff's injury and, if possible, how their combined effect alters the probability of an injury.

In this Part we offer some examples designed to suggest how courts might proceed with this very difficult competing-cause issue. However, the toxicological data currently available to us do not allow us to observe causal processes and interactions between competing risks, and a complete understanding of how risks

---

180. For a general discussion of biomarkers, see Gold, *supra* note 91.

181. No. Civ. A. 97C-06-194 CH, 2005 WL 1952859 (Del. Super. Ct. June 23, 2005), *aff'd*, 906 A.2d 787 (Del. 2006).

182. For a similar case, see Wintz *ex rel.* Wintz v. Northrop Corp., 110 F.3d 508 (7th Cir. 1997).

183. Recall one plaintiff expert argued that in some unexplained way the fungicide and the mutation acted together to cause her injury. If, in fact, the expert had evidence that the phenotype of the gene was moderated by toxic exposure, a different result might be justified.

combine and interact is currently well beyond our reach. Again, we are left with group data, but macro information about the frequencies of joint appearances in epidemiological studies that we discuss below have serious limitations. In the absence of a more complete understanding of this issue we, and the courts, must proceed with some simplifying decision heuristics based on group data, tort rules, and policy considerations.

*1. Assessing the Strength and Interaction of Competing "Causes"*

At the outset, we should note that one can imagine some easy cases where alternative causes, although present, appear to be *de minimus*. If the choice of whether one's lung cancer was the result of a lifetime of heavy smoking or by a brief encounter with a substance for which there is a significant but weak correlation with lung cancer, in most situations it should be an easy task to rule out the other substance as the specific cause of the individual's injury. The issue becomes more difficult when the other causes pose more significant threats.

Here is a simplified hypothetical example. Imagine an individual consumes two pills together, each ingested once, and suffers an injury common to ingestion of each pill. Suppose we have good data comparing the outcomes for those exposed to each pill singly to the outcomes of those exposed to neither pill. The relative risk of injury from pill one alone is 3.0. The relative risk of the same injury from pill two alone is 6.0. In what way are the individual risks of 6.0 and 3.0 relevant to the differential etiology of the individual exposed to both? The answer to this question requires us to ask some additional questions.

An initial question is how risk factors combine and interact to affect the probability of an injury. Combinations are typically examined using statistical models in which either (a) the excess ("attributable") risks or (b) the relative risks produced by each component are presumed to be unaffected by the presence or absence of other components. These presumptions imply that the total risk of a combination is respectively the sum (additive model) or the product of the contributions of its components (multiplicative model). In our example above, the relative risk compared to unexposed individuals would be $1 + (6-1) + (3-1) = 8$, (risk differences add, similar to Fig 2 below) in an additive model and $(6 \times 3) = 18$, in a multiplicative model.

In the discussion of external validity, we introduced the idea of an effect modifier: a variable that modifies the relationship between the alleged cause and plaintiff's condition. [184] Such additional factors that alter the strength of a relationship between the risk factor under investigation and a particular injury are not uncommon. "Synergism" and "antagonism" are terms used to describe circumstances in which combinations of risk factors produce considerably greater or weaker risks than their individual components would lead one to expect,[185] but the application of these terms depends on which model generates one's expectations. In our example, antagonism would exist if the joint effect is less than the sum of each separately in an additive model or less than the product of each separately in a

184. *See supra* p. 25.

185. John Darroch, *Biologic Synergism and Parallelism*, 145 AM. J. EPIDEMIOLOGY 661 (1997).

multiplicative model. A well-known example of a synergistic effect is the combined effect of asbestos exposure and smoking on the likelihood of developing lung cancer. For long-term smokers, the relative risk of developing lung cancer compared to those who have never smoked is sometimes estimated to be in the range of 10.0. For individuals substantially exposed to asbestos, the relative risk of developing lung cancer compared to nonexposed individuals is in the range of 5.0.[186] However, if one is unfortunate enough to have been exposed to asbestos and to have been a long-term smoker, the relative risk compared to those unexposed individuals who have not smoked exceeds the sum of the relative risks. One possibility is that the relationship is multiplicative, in the range of 50.0—i.e., a 49-fold risk increment.[187] In this situation, the "joint effect" of the two combined exposures is greater than the sum of the effects associated with each alone: one may view a four-fold increment (RR=5) as due to asbestos exposure, a nine-fold increment (RR=10) as due to smoking, and the remaining 49−4−9=36-fold increment as due to their joint effect. Such a relationship would be synergistic on the additive scale because the additive increase in lung cancer incidence due to either factor is substantially increased in the presence of the other, but this is not true on the multiplicative scale. Were the relationship purely multiplicative, the relative risk of either factor would be unaffected by the presence of the other. Meta-analyses now tend to indicate that the relationship is more than additive but less than fully multiplicative.[188] This means that the relative risk for nonsmokers exposed to asbestos is higher than the relative risk for smokers who are exposed to asbestos.[189] From a public-health perspective, which stresses the additive model because that often best summarizes impact upon

---

186. FAIGMAN ET AL., *supra* note 11, § 26.25.

187. *Id.*

188. *See* Peter N. Lee, *Relation Between Exposure to Asbestos and Smoking Jointly and the Risk of Lung Cancer*, 58(3) OCCUPATIONAL & ENV'T MED. 145 (2001) (data support a multiplicative model); Darren Wraith & Kerrie Mengersen, *Assessing the Combined Effect of Asbestos Exposure & Smoking on Lung Cancer: A Bayesian Approach*, 26(5) STATS. MED. 1150, 1150 (2007) (evidence supports more than an additive model and less than a multiplicative relation); Gillian Frost et al., *The Effect of Smoking on the Risk of Lung Cancer Mortality for Asbestos Workers in Great Britain (1971–2005)*, 55(3) ANNALS OCCUPATIONAL HYGIENE 239 (2011) (interaction is greater than additive but a multiplicative model could not be ruled out); Yuwadee Ngamwong et al., *Additive Synergism Between Asbestos and Smoking in Lung Cancer Risk: A Systematic Review and Meta-Analysis*, 10 PLoS ONE e0135798 (2015) (results point to a strong additive synergism for lung cancer with co-exposure of asbestos and cigarette smoking).

189. Geoffrey Berry & F.D.K. Liddell, *The Interaction of Asbestos and Smoking in Lung Cancer*, 45(5) ANNALS OCCUPATIONAL HYGIENE 341 (2001) estimates that overall nonsmokers have a relative risk of lung cancer due to asbestos that is 2.04 times that of smokers. There is also some research on the association for smoking and a number of other non-tobacco risk factors for cardiovascular disease. *See* Jay H. Lubin et al., *Synergistic and Non-Synergistic Associations for Cigarette Smoking & Non-Tobacco Risk Factors for Cardiovascular Disease Incidence in the Atherosclerosis Risk in Communities (ARIC) Study*, 19 NICOTINE & TOBACCO RSCH. 826 (2017).

The calculation of synergistic effects is fairly complex. Much of the writing on this topic has to do with synergistic effects of drugs (where synergism is often a desirable trait). *See, e.g.*, Ting-Chao Chou, *Drug Combination Studies and Their Synergy Quantification Using the Chou-Talalay Method*, 70 CANCER RSCH. 440 (2010).

a population, such a relationship is considered synergistic. However, from an etiologic perspective, where multiplicative models are most commonly employed, the effect of each factor is diminished in the presence of the other.[190]

Noel Weiss provides an example of this type of situation in his discussion of the relationship between estrogen use and endometrial cancer in the situation where an individual is or is not obese. Obesity is a risk factor for this form of cancer that is unrelated to the use of estrogen. He presents the following hypothetical data.

| | | Postmenopausal Estrogen Use | | | |
|---|---|---|---|---|---|
| | | Never | Current ≥ 5 years | Risk Difference | Relative Risk (Risk Ratio) |
| Obesity | No | 50 | 200 | 150 | 4 |
| | Yes | 150 | 300 | 150 | 2 |
| | Risk Difference | 100 | 100 | | |
| | Relative Risk (Risk Ratio) | 3 | 1.5 | | |

*Figure 2: Annual Incidence of Endometrial Cancer per 100,000 Women, by Obesity and History of Estrogen Use*[191]

These data illustrate an additive model, meaning that (a) the increases in cancer incidence from both obesity and estrogen use do not depend on whether the other risk factor is present or absent, and thus (b) the increase in incidence among women with both risk factors from the incidence among those with neither (here, 300 − 50 = 250 per 100,000 women) is the sum of the increases associated with each individually in the absence of the other (here, 150 + 100 = 250, the same result). In contrast, in a multiplicative model, the ratio of the incidences in those with to those without each risk factor is unaffected by the presence or absence of the other, and the risk ratio among those with both relative to those with neither is the product of risk ratios of each individually (here, 4 × 3 = 12). Thus, if the same risk factors combined multiplicatively, in contrast to the table above, one would expect an incidence rate of 12 × 50 = 600 cases per 100,000 among women with both, twice the expected incidence in the additive model and an increment of (600−50) =

190. INT'L AGENCY FOR RSCH. ON CANCER, WORLD HEALTH ORG., IARC MONOGRAPHS ON THE EVALUATION OF CARCINOGENIC RISKS TO HUMANS, NO. 83, TOBACCO SMOKE AND INVOLUNTARY SMOKING 913–19 (2004).

191. Adapted from FAIGMAN ET AL., *supra* note 11, § 23.46.

550 or, equivalently, 4 (RR Estrogen) × 3 (RR Obesity) − 1 = 11 times the baseline risk of 50 for patients who are neither obese nor estrogen users.

For the additive model to apply, as in the table above, the presence of each factor must attenuate the relative risk of the other. In the hypothetical data this attenuation is by exactly half, from 4.0 to 2.0 for obesity and 3.0 to 1.5 for estrogen. If nothing else, this example and the asbestos–smoking example indicate that the use of general-causation epidemiologic data to resolve specific-causation issues can become complicated very quickly. The examples raise several specific-causation issues. When an individual has another risk factor that independently causes the same disease, what role should the joint effect of these factors play in assessing specific causation? There is no one clearly correct answer. One aspect of the problem emerges when comparing the asbestos–smoking and the hypothetical estrogen–obesity examples. In the former, substantially more than half the cases among those exposed to both are apparently due to the joint exposure, and one may conclude that for anyone jointly exposed who develops lung cancer it is more likely than not that their illness is due to the joint exposure (36/50).[192] However, this is not the case in the estrogen–obesity example. The "joint risk" is not the more-likely-than-not cause of the injury. Therefore, one approach would be to look at each independent risk separately.

If an obese individual experiences a doubling of the risk from taking estrogen, should this resolve the specific-causation question in favor of the plaintiff?[193] If we compare the individual with other obese individuals who did not take estrogen, the answer is yes. Taking estrogen doubled risk. But one could argue that obesity is a greater risk factor than estrogen use. How should we view obesity? A variation on the above example that is similar to our earlier two-pill example in which obesity is replaced by another drug, Drug X, helps to address this question.

Here, Drug X is clearly a competing cause of endometrial cancer. From a specific-causation perspective and looking at each risk factor alone, it is more likely that estrogen use caused the injury of a person that took both drugs, and thus a lawsuit against a Drug X defendant should fail. Should we treat obesity and Drug X similarly, that is, treat both as competing causes? Should we treat all risk factors we can calculate as competing causes? As a matter of risk alone, this is an appropriate approach. However, one may ask whether legally we should treat every background risk such as obesity as a competing cause. The answer to this question is as much a matter of policy as it is of risk. We discuss cases that deal with the issue below.[194]

---

192. That is, of the total 49-fold increment, 9-fold is due to smoking and 4-fold to asbestos exposure. The remaining 36-fold increment is attributable to their joint occurrence.

193. Again, we emphasize that this discussion assumes that there are not significant internal- and external-validity problems with the research in question.

194. Many of these cases wrestle with the even more complex question of how to deal with situations where the individual has multiple foreground and background risk factors. The above example presents a multiple-risk-factor situation if we reintroduce obesity along with Drug X and estrogen as a third risk.

| | | Postmenopausal Estrogen Use | | | |
|---|---|---|---|---|---|
| | | Never | Current ≥ 5 years | Risk Difference | Relative Risk (Risk Ratio) |
| Drug X | Never | 50 | 200 | 150 | 4 |
| | Current ≥ 5 years | 150 | 300 | 150 | 2 |
| | Risk Difference | 100 | 100 | | |
| | Relative Risk (Risk Ratio) | 3 | 1.5 | | |

*Figure 3: Annual Incidence of Endometrial Cancer Per 100,000 Women, by Exposure to Drug X and History of Postmenopausal Estrogen Usage*[195]

Assuming that a court does choose to treat either obesity or Drug X as a potential cause of the plaintiff's injury, how does this affect the specific-causation inquiry? Absent a legal rule to the contrary, if the individual plaintiff must prove that it is more likely than not that a risk factor is the cause of her injury, she will fail if there is one or more competing-cause risk factors that have a stronger relationship with the effect in question and where the joint effect of the risks is not greater than each risk's independent effect.[196]

In the Drug X and obesity examples, if they are viewed as competing causes, the plaintiff cannot show that it is more likely than not that estrogen is the cause of the cancer. Tort law, however, is somewhat ambiguous as to how we should treat this case. Section 26 of the Restatement (Third) of Torts: Liability for Physical and Emotional Harms adopts a but-for test of causation. It states, "Tortious conduct must be a factual cause of harm for liability to be imposed. Conduct is a factual cause of harm when the harm would not have occurred absent the conduct. Tortious conduct may also be a factual cause of harm under § 27."[197]

Comment *e* of § 26 contains the following passage that addresses the types of situations presented in the preceding hypotheticals:

> In some cases, two causal sets may exist, one or the other of which was the cause of harm. Thus, for example, in a case in which the plaintiff claims that a vaccination caused subsequent seizures, and the defendant claims that the seizures were caused not by the

195. Adapted from FAIGMAN ET AL., *supra* note 11, § 23.46.

196. In some cases, courts have admitted testimony and noted that the risk factor under consideration has a stronger relationship with the effect than other known risk factors. Cooper v. Takeda Pharms. Am., Inc., 191 Cal. Rptr. 3d 67, 79 (Ct. App. 2015).

197. RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYSICAL AND EMOTIONAL HARM § 26 (AM. L. INST. 2010).

> vaccination, but by a preexisting traumatic injury to the plaintiff, the causal set including the vaccination and the causal set including the traumatic injury are such alternative causes. If sufficient evidence to support each of these causal sets is introduced, the factfinder will have to determine which one is better supported by the evidence. On the other hand, if the evidence revealed that a traumatic injury and a vaccination can interact and cause seizures, then the vaccination and the trauma may each be a factual cause (both elements of the causal set) of the plaintiff's seizures. Section 27 addresses the unusual case where each of the causal sets is independently capable of causing the harm and would have been a but-for cause if the other causal set had not existed. This phenomenon is labeled in § 27 as "multiple sufficient causes." Section 27 provides a standard for identifying additional factual causes in this circumstance.

Section 27 states, "If multiple acts occur, each of which under § 26 alone would have been a factual cause of the physical harm at the same time in the absence of the other act(s), each act is regarded as a factual cause of the harm."[198] Does this section apply to situations such as that described in the obesity or the Drug X hypothetical?

A narrow reading of § 27 and the comments accompanying the section suggest that the answer is no. Comment *e* to § 27 distinguishes between multiple sufficient causes (e.g., the two-fire situation) and alternative causes, as follows:

> In some cases, a defendant may contend that the acts of another were the cause of the plaintiff's harm and thus that defendant's tortious conduct was not a cause of the plaintiff's harm. Whether that claim implicates the rule in this Section depends on whether the other forces were operating and sufficient to cause the harm contemporaneously with the defendant's tortious conduct or, alternatively, were the factual cause of the harm *instead of* the defendant's tortious conduct. If the evidence supports the former finding, then this Section is applicable. If the evidence supports the latter finding, then the applicable standard for factual causation is that stated in § 26.[199]

---

198. *Id.* § 27. The origin of the rule in § 27 is the problem presented in what are commonly called the "two fire" cases. A forest fire negligently started by person A to the northeast of the plaintiff's house merges with a fire negligently started by person B to the northwest of the plaintiff's house. The combined fire burns down the plaintiff's house. Under the generally applied but-for causation, but for the defendant's negligent act, the plaintiff would not have been injured, and the plaintiff must prove this by a preponderance of the evidence (more likely than not). So, the plaintiff sues A. But A says you cannot prove but-for causation because even if my fire had never existed, B's fire would have burned down your house, and in fact this is true. And of course, a suit against B produces the same answer in reverse. What to do? The law's answer is to resolve the causal problem with the § 27 exception to the § 26 rule.

199. *Id.* § 27 cmt. e.

The contemporaneous requirement and the "instead of" in this comment may be understood to restrict the application of § 27 to situations like the two-fire scenario in footnote 198. Given this interpretation, most courts would refuse to apply § 27 to most specific-causation situations. However, an illustration immediately following the language quoted above creates some uncertainty as to which toxic-tort situations might fall within the scope of § 27. It reads:

> Trent is the guardian ad litem and father of Lakeesha, an infant born with a birth defect. Trent sues Pharmco, a pharmaceutical company, alleging both that Pharmco's drug caused Lakeesha's birth defect and that Pharmco was negligent for its failure to warn that its drug was teratogenic. Trent introduces sufficient evidence for the factfinder to find that Pharmco's failure to warn was negligent, that the drug was a cause of Lakeesha's birth defect, and that an adequate warning would have prevented the birth defect. Pharmco contends that its drug did not cause Lakeesha's birth defect. Rather, Pharmco contends, Lakeesha's birth defect was caused by a genetic condition wholly independent of Pharmco's drug. Pharmco introduces sufficient evidence in support of its claims. The factfinder must determine if the drug, absent Lakeesha's genetic condition, would have caused the birth defect. The factfinder must also determine if, absent the drug, Lakeesha's genetic condition would have caused the birth defect. If the factfinder determines that either the drug or the genetic condition would have, in the absence of the other, caused Lakeesha's birth defect at the same time then each is a factual cause pursuant to this Section. If the factfinder determines that either the drug or the genetic condition played no role in the birth defect, then the other's causal status is determined under the but-for standard of § 26.[200]

This illustration suggests that there could be some two-cause situations where § 27 does apply. The outcome turns, it appears, on how one chooses to understand the terms "at the same time" and "wholly independent of." The "wholly independent of" language presents a particularly high barrier if we understand it to require that two substances, such as exist in the Drug X example, must produce the same injury by two completely independent mechanistic pathways once they enter the individual's body. Unfortunately, it is currently beyond our ability to map these pathways, but what we do know about some diseases, such as cancer, is that they involve multiple mutations affecting a number of genes.[201]

The adoption of a § 27 position would mean that in cases where the relative risk of an injury exceeds 2.0 among individuals sharing the plaintiff's characteristics, the way most courts describe the differential-etiology task is incorrect. Courts describe the process of differential etiology as an effort to "rule out" all causes other than the putative cause asserted by the plaintiff. But under § 27, if the putative cause has a relative risk greater than 2.0 (with valid research

200. *Id.* § 27, cmt. e, illus. 2.
201. *See* Gold, *supra* note 91.

literature underlying it), it doesn't matter if other causes are present, even if they have high relative risks. When multiple sufficient causes are present in the plaintiff's case, therefore, the ultimate issue is not whether the putative cause more likely than not was the cause-in-fact of the condition. Instead, the test asks whether the substance in question more likely than not is sufficient to cause the condition, even in the presence of other causes that might have been equally likely, or more likely, to be the cause-in-fact. As long as the relative risk is greater than 2.0, the research meets adequate internal-validity standards, and the plaintiff's condition is represented by the research (i.e., external validity), the expert testimony would be admissible without more.[202]

We do not know of any opinions that explicitly adopt this position. Perhaps, however, this is not surprising because we have not uncovered any opinions that refuse to adopt the § 27 position where it is clear that the relative risk of the cause in question is unarguably greater than 2.0, but there are competing causes that present an even greater risk. The obvious place to find such cases would be plaintiffs with lung cancer suing a defendant that had exposed them to a substantial dose of asbestos where the record shows that the plaintiff was a long-time smoker. However, we are aware of no cases where the court has excluded asbestos expert testimony on these grounds.[203] The absence of such cases suggests that when courts are presented with two competing causes, each of which by itself poses a relative risk greater than 2.0, they may be reluctant to absolve the cause posing the lesser risk on specific-causation grounds.[204]

The significance of this observation should not be underestimated. The gravamen of differential etiology is to identify *the* cause-in-fact of the plaintiff's

---

202. Similar ambiguity exists in jurisdictions that continue to use a "substantial factor" test for causation. Guinn v. AstraZeneca Pharms. LP, 598 F. Supp. 2d 1239, 1246 (M.D. Fla. 2009). Take the following example from Florida:

> Florida's concurring cause rules are, at best, unsettled. In particular, it is unclear whether, as [*Stahl v. Metropolitan Dade County*, 438 So.2d 14, 19 (Fla. Dist. Ct. App. 1983)] states, the "substantial factor" test applies "only in those concurring cause cases where each of the said concurring causes could have alone produced the plaintiff's injury," or whether, as Florida Standard Jury Instruction 5.1(b) and the preexisting condition cases cited above can be read to suggest, causation is satisfied where the plaintiff establishes that the defendant's negligence substantially contributed to the plaintiff's injury, even though such negligence was one of perhaps many concurring causes that acted in combination to cause the injury.

203. There are cases where a court has permitted tobacco defendants to introduce evidence of plaintiff's asbestos exposure. *See, e.g.*, Tompkin v. Philip Morris USA, Inc., 362 F.3d 882 (6th Cir. 2004). Another place one might find such rulings would be in cases where the plaintiff simultaneously sued asbestos and tobacco defendants. Surprisingly, such cases are quite rare, sometimes because courts have held that such joinder is impermissible. *See* R.J. Reynolds Tobacco Co. v. Stidham, 141 A.3d 1, 12–14 (Md. 2016) (overruling cases that hold such joinder is never permissible),

204. In the asbestos–tobacco situation, allowing such testimony could be justified by the argument presented above that the joint effect of asbestos and tobacco exposure creates a risk greater than the risk posed by either exposure in the absence of the other. However, no court has ever actually adopted this approach.

injury by ruling in the plaintiff's claimed cause and ruling out alternative causes. If, for instance, the plaintiff's claimed Cause X has a relative risk of 2.1, but the plaintiff was exposed to Cause Y with a relative risk of 20, then other things being equal Y would appear to be more incriminated as the actual cause-in-fact of the plaintiff's condition. However, if a court allows proof of Cause X in such cases to go to the jury—as appears to be the practice in some areas—it is eschewing differential etiology as an evidentiary threshold matter when the plaintiff's proof is adequate to conclude that his or her claimed cause, when considered in isolation, would be sufficient to double the risk of injury, even if it is not more likely than not the actual cause of the injury in the case at hand. Hence, judging by what appears to be the current practice in asbestos–smoking cases, and perhaps contrary to conventional understanding, differential etiology is only relevant when the plaintiff's proof falls below a relative risk of 2.0, is based on less-than-sound research literature, or no relative risk is available because of the absence of epidemiological research. Cases where there are multiple risk factors—and the risk being litigated does not have relative risks greater than 2.0—generally parallel the *In re Lipitor* opinion mentioned at the beginning of this Article.[205] Recall that the plaintiff, Ms. Jones, alleged that the drug, taken to control cholesterol levels, caused her diabetes. The court excluded her specific-causation expert because studies indicate the relative risk associated with taking the drug is less than 2.0. The court also noted that the plaintiff was overweight with a body mass index of nearly 28 at the time she was diagnosed with diabetes.[206] In addition, she suffered from hypertension, and her family had some history of diabetes. All of these are risk factors for diabetes[207] as is age by itself.[208] Several of them appear to have a relative risk associated with contracting diabetes substantially greater than the relative risk associated with taking Lipitor.[209] In this situation, and if the objective is to find the cause of the plaintiff's injury to the exclusion of other possible causes, it is difficult—if not impossible—for the plaintiff to demonstrate that Lipitor is more likely than not the cause of her

---

205. *In re* Lipitor (Atorvastatin Calcium) Mktg., Sales Pracs., & Prods. Liab. Litig., 150 F. Supp. 3d 644, 649–51 (D.S.C. 2015), *aff'd*, 892 F.3d 624 (4th Cir. 2018).

206. *Id.* at 648.

207. The court refers to an American Journal of Epidemiology article based on the well-regarded Nurses' Health Study that found women with a body mass index ("BMI") of 25 to 26.9 had more than a five-fold increased risk of diabetes compared to women with a body mass index below 22.14. Yan Zheng et al., *Group-Based Trajectory of Body Shape from Ages 5 to 55 Years and Cardiometabolic Disease Risk in 2 U.S. Cohorts*, 186 AM. J. EPIDEMIOLOGY 1246 (2017). In addition, adult weight gain independent of BMI increases the risk of diabetes. The Nurses' Health Study found a weight gain in the range experienced by the plaintiff, adjusted for BMI, was associated a two-fold increase in the risk of diabetes. *Id.*

208. Age as a risk factor puts a fine point on the question of whether background causes are treated as competing causes. Age, of course, is a surrogate for other biological processes that change as we age. It is, therefore, a risk factor for many ailments.

209. Speaking of plaintiff's expert, the court noted:

> [S]he has quantified the relative risk due to Lipitor as 1.6 and has admitted that studies have shown the relative risk of other factors were much higher. For example, Ms. Hempstead's BMI gave her a relative risk over 5.0, her total adult weight gain gave her a relative risk of 12.0, and her hypertension gave her a relative risk of 1.76.

*In re Lipitor*, 150 F. Supp. 3d at 656–57.

diabetes. But we should recognize that at present this task is beyond anyone. What is clear is that cumulatively other risk factors far outweigh the apparent effect of Lipitor. When this is the situation, many courts are prepared to exclude plaintiff's proffered expert testimony.

Sometimes the existence of other known potential causes offers another approach to dealing with situations when epidemiologic studies suggest general causation with a relative risk less than 2.0. This is the case because the estimated relative risk between a particular factor and an injury typically is "marginal;" that is, it is averaged over populations among which individual reactions to exposures may vary, and there may be evidence distinguishing the plaintiff's sensitivity to the exposure from an average relative risk reported in the epidemiologic studies.[210] Again, when competing causes act as effect modifiers, they may prove useful in the specific-causation analysis.

Effect modifiers may come in several forms. Sometimes, the addition of another factor may work to the benefit of the plaintiff's specific-causation argument. For example, the plaintiff may be more sensitive to a specific exposure.[211] Biomarkers of greater sensitivity may help an individual to show that it is more likely than not that a substance caused her injury.[212] As the NAT2 gene variation and breast-cancer example indicates, the relative risk of an injury following an exposure may vary depending on an individual's specific genetic makeup. The NAT2 example and the Weiss and Drug X hypotheticals underline an intuitive point that whenever there is an ability to specify subgroups within a study and then calculate relative risks for a subgroup of at least moderate sample size, this leads to a more valid estimate of the risk of an exposure —in the sense that it is less subject to external-validity concerns—for a given individual who can be placed in the relevant subgroup. Interestingly, and more subtly, this may also be so even when the individual cannot be placed in a specific subgroup with certainty.[213] Exposure

---

210. For a case discussing this possibility, see *Estate of George v. Vermont League of Cities & Towns*, 993 A.2d 367 (Vt. 2010).

211. Some variation is the result of differential absorption. For example, iron deficiency, which is more likely among disadvantaged children, increases the rate of absorption of ingested lead. The rate of metabolic processes also may vary across individuals. Slower metabolic processes among older people may alter the rate of elimination of a toxin.

212. Earlier we cited *Kirk v. Schaeffler Group USA, Inc.*, 887 F.3d 376, 391 (8th Cir. 2018) for a case where the plaintiff's expert testified to the plaintiff's genetic predisposition that caused her to develop autoimmune hepatitis ("AIH") as a result of exposure to defendant's trichloroethylene.

213. A. Philip Dawid, *The Role of Scientific and Statistical Evidence in Assessing Causality*, *in* PERSPECTIVES ON CAUSATION 133–47 (Richard Goldberg ed., 2011). The point is made by the following example (slightly edited from the original):

> Suppose that scientific research has identified a gene with two variants, G and N, each present in 50% of the population. It has been found that, among those having variant G, the headache recovery rate (within 30 minutes) if untreated is 24%, but because of an interaction with the action of aspirin, such an individual will never recover from a headache if treated with aspirin. As for those with variant N, they never recover if left untreated but have a 60% chance of recovery if they take aspirin. Over the whole population (ignoring genetic status), the recovery rate is 30% with aspirin and 12% without aspirin. Using only

biomarkers may also play a role. For example, the absence of a high-validity biomarker of exposure to a rival cause of an individual's illness offers some evidence excluding this alternative cause.[214]

As useful as it may be in some situations, the subgroup solution comes with its own limitations. Subgroup data are always less plentiful, and hence estimated relative risks in subgroups are generally substantially less precise than overall estimates. Thus, even though subgroup estimates may be less biased than overall (crude) estimates, in any particular instance they may well depart further from the truth due to increased random variation. This sample-size limitation quickly becomes decisive when we contemplate multiple competing causes as in the Lipitor

---

these overall figures, given that I took the aspirin and recovered, the lower bound on the probability of causation ("PC") is 60%.

However, with the additional scientific knowledge of the gene, given that I did recover after taking aspirin, it can be inferred that I cannot be of type G, so I must be N; and it then follows that I would certainly not have recovered if I had not taken the aspirin. So, with this extra scientific knowledge we can deduce PC = 1 – a major improvement on (although of course consistent with) the "black-box" conclusion PC $\geq$ 60%.)

In more typical cases, such scientific knowledge of background attributes will not fully determine any outcome. Nevertheless, even for a case where we do not know the background attributes, we can still use that scientific knowledge to obtain an improvement in the lower bound for PC (though not now to 1). A formula for this is given in the cited article.

214. Stout & Valberg, *supra* note 62, at 802. The probative value of the absence of a marker depends on the marker's sensitivity and specificity. In this context, sensitivity measures how often the marker is present in individuals known to have been exposed. Specificity measures how often the marker is absent in persons known not to have been exposed.

One must not fall into the trap of concluding that the existence of a biomarker of exposure can resolve the specific-causation question. The existence of a biomarker does not rule out the possibility that some other substance or exposure caused the individual's injury. Moreover, the markers we do have are rarely very specific or sensitive. Most toxins or their metabolites are thought to produce illness by multiple pathways, and many pathways are thought to be shared by multiple toxins. Moreover, epigenetic effects—that is, effects of factors that regulate the expression of genes—further complicate the interpretation of a genetic biomarker.

To illustrate the difficulty in finding biomarkers that are highly specific and highly sensitive, consider again the connection between benzene exposure and leukemia. Studies have found chromosomal aberrations that sometimes occur in leukemia patients with known occupational benzene exposure, but they also occur in leukemia patients in the control groups of these studies. Despite the progress in chromosomal and genetic study of leukemia in people exposed to benzene, the search for a signature biomarker of exposure continues. And even if we do discover such a marker, this, by itself does not rule out the possibility that some other exposure or some genetic factor caused the individual's illness. As is true with epidemiological associations, the level of reliance one should place on biomarker research turns on questions of validity, consistency, specificity, coherence, and biological plausibility. Some greater headway has been made with respect to showing a genetic rather than a toxic source of an injury. This was the case in *Deribeaux ex rel. Deribeaux v. Secretary of Health & Human Services*, 717 F.3d 1363 (Fed. Cir. 2013), where the government was able to show that the claimant's SCN1A gene mutation rather than the Diphtheria–Tetanusacellular Pertussis ("DTaP") vaccine was the sole substantial cause of Deribeaux's seizure disorder and developmental delays.

case, where other risk factors included age, weight, hypertension, and a family history of diabetes. Moreover, in complex situations such as this we may know little more than that the absolute risk for someone with all these factors is greater than the absolute risk for those individuals with only one risk factor.[215]

Thus, to date, the analysis of how competing causes influence each other has been relatively unsophisticated. It remains to be seen whether advances in causal-inference methods may at some time in the future enhance our ability to address these complex relationships.

*2. The Problem of Idiopathic Causes*

Cases where a large percentage of injuries are the result of unknown causes present a similar but even more difficult admissibility problem for plaintiffs.[216] All diseases have some unknown—or in medical terms, "idiopathic"—causes, if for no other reason than a failure to discover an exposure to a known cause. This fact seems to be ignored in most situations. Below we discuss some cases in which a substantial majority of injuries of the relevant type occur due to unknown causes and where courts have excluded expert differential-diagnosis opinions on these grounds.

The rather well-known case of *Milward v. Acuity Specialty Products Group, Inc.* ultimately turned on this issue. The plaintiff had prevailed on the general-causation issue of whether benzene exposure can cause the plaintiff's rare form of Acute Myeloid Leukemia ("AML").[217] On remand, the trial court noted that between 70 and 80 percent of AML cases are idiopathic. In light of this fact, it made

---

215. There is one more point about the *Lipitor* opinion worth mentioning. The plaintiff's expert opined that even if Lipitor did not "cause" the plaintiff's injury in the "but for" sense—that if the plaintiff had not taken Lipitor she would never have become diabetic—it caused the plaintiff to develop the disease earlier than would otherwise have occurred. Regardless of the merits of the argument on these facts, it is a version of a "promotion" argument. The same argument was made in *Haller v. AstraZeneca Pharmaceuticals LP*, 598 F. Supp. 2d 1271 (M.D. Fla. 2009). The plaintiff's expert testified, "Now, ask me to take away Seroquel, well, it would still have happened, but it may not have happened until many years later, ma'am." *Id.* at 1288–89. A similar argument also appeared in the *Joiner* case.

If in fact the promotion argument is correct, it poses additional problems of specific causation. As Greenland notes, if a promotion (or, as he calls it, an "acceleration") argument is correct, then the relative risk in a study underestimates the probability that exposure accelerated the occurrence of an injury. *See* Sander Greenland, *Relation of Probability of Causation to Relative Risk and Doubling Dose: A Methodologic Error That Has Become a Social Problem*, 89 AM. J. PUB. HEALTH 1166 (1999). Other statistical methods must be brought to bear to analyze this argument. Survival models explicitly model acceleration (i.e., promotion) or retardation by expanding or shrinking the time scale. *See* DAVID W. HOSMER, JR. ET AL., APPLIED SURVIVAL ANALYSIS: REGRESSION MODELING OF TIME-TO-EVENT DATA (2d ed. 2008); HANDBOOK OF SURVIVAL ANALYSIS (John P. Klein et al. eds., 2013).

216. In this context, idiopathic risks are risks we simply do not know and also risks that we know to exist but do not understand. An example of the latter is a general knowledge that one's genetic makeup influences the probability that exposure to a "cause" will produce an effect in an individual but without any understanding of how this occurs.

217. 639 F.3d 11 (1st Cir. 2011).

the following observation about the proffered testimony of the plaintiff's specific-causation expert:

> Here, even if Butler could rule out smoking and obesity as probable causes, the differential diagnosis analysis provides little information. When a disease has a discrete set of causes, eliminating some number of them significantly raises the probability that the remaining option or options were the cause-in-fact of the disease. Restatement (Third) of Torts: Phys. & Emot. Harm § 28, cmt. c (2010) ("The underlying premise [of differential etiology] is that each of the [ ] known causes is independently responsible for some proportion of the disease in a given population. Eliminating one or more of these as a possible cause for a specific plaintiff's disease increases the probability that the agent in question was responsible for that plaintiff's disease."). The same cannot be said when eliminating a few possible causes leaves not only fewer possible causes but also a high probability that a cause cannot be identified. ("When the causes of a disease are largely unknown . . . differential etiology is of little assistance."). Butler cannot establish specific causation in this context using a differential diagnosis approach.[218]

Other cases reaching a similar result include *Doe v. Ortho-Clinical Diagnostics, Inc.*,[219] *Hall v. Conoco Inc.*,[220] *Henricksen v. ConocoPhillips Co.*,[221] *Kilpatrick v. Breg, Inc.*,[222] and *Perry v. Novartis Pharmaceuticals Corp.*[223]

Two things are worth noting about these cases. First, in none of these cases is there group data suggesting even a doubling of risk of the injury in question among those exposed to the substance.[224] In most of these cases (*Doe*, *Hall*, and *Kilpatrick*), there is in fact little or no data on general causation. In *Hall* and *Milward*, the plaintiff had very little data on dosage, which apparently was low.[225] Second, in the

---

218. Milward v. Acuity Specialty Prods. Grp., Inc., 969 F. Supp. 2d 101, 109 (D. Mass. 2013). Even in the face of this high percentage of idiopathic illnesses, the court stated that it would have admitted the expert's testimony had there been data on a threshold of harmful exposure, but apparently the expert simply argued that there is no safe threshold. The district court opinion was affirmed by *Milward v. Rust-Oleum Corp.*, 820 F.3d 469, 471, 477 (1st Cir. 2016).

219. 440 F. Supp. 2d 465, 477–78 (M.D.N.C. 2006).

220. 886 F.3d 1308, 1314 (10th Cir. 2018).

221. 605 F. Supp. 2d 1142, 1162 (E.D. Wash. 2009).

222. 613 F.3d 1329, 1342–43 (11th Cir. 2010).

223. 564 F. Supp. 2d 452, 471 (E.D. Pa. 2008).

224. Thus, even were a court willing to adopt a multiple causal set analysis as discussed above, these cases would not benefit from such an analysis.

225. The relative risks associated with low doses of benzene and Acute Myeloid Leukemia tend to be well below 2.0. *See* Abdul Khalade et al., *Exposure to Benzene at Work and the Risk of Leukemia: A Systematic Review and Meta-Analysis*, 9 ENV'T HEALTH 31 (2010); Martha S. Linet et al., *Benzene Exposure Response and Risk of Myeloid Neoplasms in Chinese Workers: A Multicenter Case-Cohort Study*, 111(5) J. Nat'l Cancer Inst. 465 (2019).

majority of known cases, the substantial majority of injuries are attributed to idiopathic causes. Therefore, there is no way for the plaintiff to differentiate herself from other individuals. A subgroup analysis is rarely possible. In this regard, *Perry* is a particularly interesting case. In *Perry*, the plaintiffs claimed their child's T-cell lymphoblastic lymphoma (T-LBL) (a type of non-Hodgkin lymphoma) was caused by her use of Elidel, a prescription cream used to treat eczema. The court excluded plaintiff's experts in part because they could not rule out idiopathic causes of the disease.

The opinion contains a discussion of what an expert must do in the circumstance where most occurrences of a disease are from unknown causes, as is the case with T-LBL. Pimecrolimus is the active ingredient in Elidel and is one of a class of drugs known as calcineurin inhibitors. Calcineurin inhibitors are known to inhibit immune-system function. Other calcineurin inhibitors are used in immunosuppressive therapy to prevent rejection after organ transplants and have been associated with increased incidence of post-transplant lymphoproliferative disorder, an illness similar to non-Hodgkin lymphoma. All of this strengthens the plaintiff's general-causation argument by presenting evidence that tends to refute the existence of a merely idiopathic origin.

When plaintiff's experts were questioned about how they excluded "no known cause" in the child's illness, they simply repeated the existence of a known risk factor, primecrolimus. The court responded: "Standing alone, the presence of a known risk factor is not a sufficient basis for ruling out idiopathic origin in a particular case, particularly where most cases of the disease have no known cause."[226]

However, the trial judge did not leave the matter here. Rather he went on to make the following comment:

> This is not to say that where most diagnoses of a disease are idiopathic it is impossible to prove specific causation. But in those cases, analysis beyond a differential diagnosis will likely be required. Here, for example, because lymphoma caused by immunosuppressant drugs is well-understood, Drs. Smith and Kolb could have compared the presentation of Andreas Perry's symptoms with those common in post-transplant lymphoma cases. Doing so, however, would not have served plaintiffs' purposes.[227]

The court explained this last sentence by noting that the post-transplant cancers have a history consistent with B-cell origin, whereas the child's lymphoma had a T-cell origin. Moreover, the post-transplant lymphomas occur, on average, 4.2 years after the transplant, with the earliest-known presentation occurring 1.7 years after transplant. The disease in this case was detected less than seven months after the child's first exposure to Elidel. These facts make the district court's task somewhat easier, as it would appear the features of this child's cancer were

---

226. *Perry*, 564 F. Supp. 2d at 470.
227. *Id.*

dissimilar to those of other cancers accepted as caused by immunosuppression.[228] However, if the illness had a B-cell origin and had occurred at the appropriate time after the initiation of Elidel, the plaintiff's case on specific causation would have been much stronger. In terms of Bradford Hill indicia, such mechanistic evidence can assist a plaintiff by providing evidence of specificity, plausibility, and coherence.[229]

Unfortunately, for the other cases listed above there is no known distinction between forms of a disease, and a court must decide if the simple failure to exclude idiopathic causes is reason for exclusion of expert testimony when in fact there is no way to exclude (or prove) idiopathic causes. In the absence of some toxicological or genetic data that can more particularly describe an injury, the idiopathic-injury cases present unique difficulties. By definition, there are no known risks we might control for by a more refined analysis of group data.[230]

Given this state of affairs, perhaps it is reasonable for courts to disregard idiopathic causes in those cases where idiopathic causes comprise a relatively small percent of all injuries.[231]

*3. Lack of Human, Especially Epidemiological Data*

A final category of cases presenting difficult admissibility questions are those where there are no epidemiologic, clinical-trial, or other group-level human data.[232] Cases where there are no relevant human-based data include injuries that are the result of a sporadic accident,[233] where ethical considerations bar research, or

228. The task was also made easier because the plaintiff faced a significant dose (i.e., general causation) problem.

229. The lack of mechanism evidence may have the opposite effect. *See* Guinn v. AstraZeneca Pharms. LP, 598 F. Supp. 2d 1239, 1246 (M.D. Fla. 2009).

230. *See* Faigman and Mnookin, *supra* note 37, at 628 for a discussion of a case that apparently ignored the fact that the plaintiff's injury resulted from unknown causes approximately three-fourths of the time.

231. It is difficult to set a specific percentage of idiopathic causes that defeat a specific-causation analysis. Suppose idiopathic cases account for 40%, cause A for 35%, and cause B for 25% of cases of a disease. Neglecting idiopathic causes fundamentally alters the probability that A is the cause from 35% to 58.33%. However, the effect of disregarding idiopathic causes depends in part on the relative importance of other competing causes. Inevitably, setting such a percentage is partly a question of policy. At least one court has simply set aside the whole issue of idiopathic causes, arguing that "[m]edicine partakes of art as well as science" and that "district courts should typically admit specific-causation opinions that lean strongly toward the 'art' side of the spectrum." *In re* Roundup Prods. Liab. Litig., 358 F. Supp. 3d 956, 959 (N.D. Cal. 2019). This, of course, is fundamentally a retreat to a *Frye*-like rule.

232. We should note at the outset that this does not mean there are no human data. Depending on the case, the plaintiff may have a biomarker of exposure or other individual-level evidence. This category is intended to include only those cases where there are no human-based group data on causation.

233. Gerald W. Boston, *A Mass-Exposure Model of Toxic Causation: The Content of Scientific Proof and the Regulatory Experience*, 18 COLUM. J. ENV'T L. 181, 188 (1993). These cases point to an underlying fact that proof of causation in one person is contingent on the existence of similarly exposed individuals. The first plaintiff (or the sole plaintiff)

where the research lags the litigation.[234] Courts consistently assert that the absence of epidemiologic evidence is not inevitably fatal to a plaintiff's proof of general causation.[235] But rarely do they then permit a plaintiff to proceed to trial. As we discussed above, in part this is because in many cases where there are no epidemiological or other human data, there often are significant general-causation issues and many of these cases are disposed of on general-causation grounds.

In the absence of human data, courts are most open to an expert's general-causation testimony when it is based on animal research. Few cases permit the expert to express a general-causation opinion based solely on *in vitro* research or on case reports.[236]

When the general-causation testimony is ruled admissible, the plaintiff still must offer admissible specific-causation testimony.[237] A number of factors affect

---

experiencing an exposure may find it much more difficult to prove causation than other, later plaintiffs who have a body of epidemiology and animal research to support their claims.

234. Schott v. I-Flow Corp., 696 F. Supp. 2d 898, 905 (S.D. Ohio 2010) (conducting epidemiological studies would be unethical.)

235. *See* Ruff v. Ensign-Bickford Indus., Inc. 168 F. Supp. 2d 1271, 1281 (D. Utah 2001); *In re* Ephedra Prods. Liab. Litig., 393 F. Supp. 2d 181, 187 (S.D.N.Y. 2005); *Schott*, 696 F. Supp. 2d at 905; Jones v. Novartis Pharms. Corp., 235 F. Supp. 3d 1244, 1269 (N.D. Ala. 2017).

236. The Bradford Hill plus indicia help us to understand why. External-validity concerns alone are often sufficient. In addition, there are often specificity problems based on level of exposure (e.g., dose), similarity between types of injury (e.g., different types of cancer), and the absence of biological gradient information.

*Graham v. Playtex Products, Inc.*, 993 F. Supp. 127 (N.D.N.Y. 1998) is a case that admitted such testimony. The Graham court allowed plaintiff's experts to testify that all cotton tampons are less likely to cause toxic shock syndrome ("TSS") than tampons containing rayon based solely on their *in vitro* research. Apparently, there was not a body of epidemiological data concerning their relative safety. The plaintiff's experts themselves seem to concede, however, that once such evidence became available, it would be inappropriate to base an opinion on the *in vitro* data. *Id.* at 131.

Most courts, however, are skeptical of causation testimony based solely on such evidence. *See In re* Abilify (Aripiprazole) Prods. Liab. Litig., 299 F. Supp. 3d 1291, 1310 (N.D. Fla. 2018); *In re* Accutane Prods. Liab., 511 F. Supp. 2d, 1288, 1294–95 (M.D. Fla. 2007); Kilpatrick v. Breg, Inc., 613 F.3d 1329, 1340-44 (11th Cir. 2010). However, as with animal studies, *in vitro* data may be used to supplement other types of evidence, provided the expert explains how the *in vitro* data can be reliably extrapolated to predict a drug's effects in humans.

There are some cases that are clearly wrongly decided. *See* Bee v. Novartis Pharms. Corp., 18 F. Supp. 3d 268, 304 (E.D.N.Y. 2014) (apparently permitting an expert to offer a causation opinion based on case reports alone). Even in the absence of epidemiology, most courts would not permit a plaintiff's expert to base a causal opinion on case reports. *See also* Meister v. Med. Eng'g Corp., 267 F.3d 1123, 1129 (D.C. Cir. 2001); Wells v. SmithKlineBeecham Corp., 601 F.3d. 375, 379–80 (5th Cir. 2010).

237. In a number of cases in this category, the court rejects the plaintiff's specific-causation opinion only after previously rejecting the general-causation evidence. Although these cases may find flaws in the specific-causation opinion, i.e., it failed to rule out other causes, the ruling was superfluous once general-causation testimony was excluded. Trial courts engage in this additional analysis presumably in order to reduce the chance of a reversal

the willingness of courts to admit such testimony. The Bradford Hill indicia again offer some guidance. The strength of a relationship remains an important consideration. However, measuring strength is not quite a straightforward enterprise. It turns in part on the dosage necessary to induce an injury in another species and how close this dose is to plausible human exposure levels. Plaintiffs who have experienced substantial exposures may more easily demonstrate specific causation. Consistent injuries across several nonhuman species bolsters a causal interpretation for humans as well, and this is especially the case if toxicologic and genetic evidence suggests a similar pathway of injury across species. Such evidence is also relevant to the indicium of biological plausibility.

Specificity is another useful indicium. Recall that in the *Joiner* case, Chief Justice Rehnquist noted that the mice in an animal study suffered from a different cancer than the plaintiff, and their exposure came through a different pathway (PCBs injected into the mice versus dermally absorbed by the plaintiff). This lack of similarity argues against the validity of the animal study to prove that the plaintiff's injury was due to exposure to PCBs.

Biological-gradient evidence by itself is not particularly important, but the absence of any gradient over a broad dose range argues against causation.[238] Of course, temporal order alone may be quite persuasive when the harm nearly immediately follows the exposure.[239]

It is important, however, not to overstate the ability of this type of evidence to establish specific causation. When there is a substantial lag time between exposure and injury, when the level of the plaintiff's exposure is not especially great,

---

on appeal. *See, e.g.*, Grant v. Pharmavite, LLC, 452 F. Supp. 2d 903 (D. Neb. 2006); Cloud v. Pfizer Inc., 198 F. Supp. 2d 1118 (D. Ariz. 2001); Henricksen v. ConocoPhillips Co., 605 F. Supp. 2d 1142 (E.D. Wash. 2009).

238. Assuming there is a biological gradient (i.e., the greater the exposure, the more likely it will result in a harm), high levels of exposure suggest a stronger relationship between the substance and resulting harm. In this regard, Judge Calabresi's opinion in *Zuchowicz v. United States*, 140 F.3d 381 (2d Cir. 1998) is instructive. In *Zuchowicz*, the plaintiff's wife died from a fatal lung condition allegedly caused by the drug Danocrine. Mrs. Zuchowicz was negligently prescribed a substantial overdose of the drug, which she took daily for over a month. She continued taking the correct dosage of the medication for another two months, when, due to adverse symptoms, she was advised to cease. According to the plaintiff's experts, because of the rareness of primary pulmonary hypertension and the lack of any formal research on the effects of the drug at this dose rate, they could not point to specific research supporting their differential diagnosis that the drug caused the decedent's illness. However, they could point to other studies showing that agents such as birth-control pills, some appetite suppressants, and chemotherapy drugs caused this illness. In support of the decision to affirm the trial court's decision to admit the testimony, Judge Calabresi stated that the experts were able to provide a biologically plausible reason why the drug could cause this effect. *Id.* at 387–90. In sum, the experts met a number of the Bradford Hill indicia of causation: strength, analogy, plausibility, and temporality. For an interesting discussion of this case, see Kenneth S. Abraham, *Self-Proving Causation*, 99 VA. L. REV. 1811 (2013).

239. *See* Marcum v. Adventist Health Sys./W., 193 P.3d 1 (Or. 2008); Tedder v. Am. Railcar Indus., Inc. 739 F.3d 1104 (8th Cir. 2014).

and when other, competing causes are present, then proof of specific causation remains very problematic in the absence of any group-level evidence.

*4. Summary: Multiple Risk Factors and Individuating Circumstances*

Situations where the individual has multiple risk factors or there is a high percentage of idiopathic causes pose especially difficult specific-causation problems. In the idiopathic situation, the variables that determine injury outcomes are, by definition, poorly understood and cannot be controlled for. In situations where the individual has many risk factors, non-epidemiologic research studies that assist in determining general causation are less useful in demonstrating specific causation. Many, such as animal studies, typically control for confounders by experimental techniques that increase internal validity but often at the expense of external validity. Consequently, they provide very limited information on the relative impacts different factors may have on a plaintiff's injury. Not surprisingly, this situation often results in the exclusion of the plaintiff's specific-causation expert.[240]

The best way to get some purchase on specific causation in these cases is, once again, to search for ways in which the plaintiff can be distinguished from others. Here, once again, the Bradford Hill-specificity indicium is particularly important. If available, the absence of biomarkers of exposure may be used to eliminate other rival causes from consideration. Markers of illness may be useful in the same way.[241] Dose and level of exposure may also play a role. If a substance exhibits a biological gradient, a very large dose increases the probability that the substance in question was the cause of an injury because the larger dose increases the strength of the relationship between the substance and the injury.[242]

Temporal order may also play a differentiating role. Sometimes the individual may be "sick too soon" or "sick too late" for the substance in question to have caused her injury.[243] But temporal order may be useful in reducing the probability of other causes as well.

These far-from-exhaustive examples are intended to emphasize the many ways that the Bradford Hill plus guidelines may assist us in unraveling the thorny question of specific causation. While none of these approaches can resolve all specific-causation issues, developing a coherent strategy for producing differentiating evidence is essential if courts are to avoid falling back into a *Frye*-like tendency to admit expert specific-causation testimony in limited evidence situations based solely on the *ipse dixit* of the expert.

---

240. *See* Kolesar v. United Agri Prods., Inc., 246 Fed. App'x 977, 981 (6th Cir. 2007); Williams v. Mosaic Fertilizer, LLC, 889 F.3d 1239, 1248–49 (11th Cir. 2018).

241. Recall the judge's discussion in *Perry v. Novartis Pharmaceuticals Corp.*, 564 F. Supp. 2d 452 (E.D. Pa. 2008). If the plaintiff's lymphoma had a B-cell origin, this would have bolstered the probability that his disease was associated with his eczema medication. The fact that it did not increased the probability of some other unknown cause.

242. Recall that this indicium played a role in *Zuchowicz*, 140 F.3d 381.

243. The *Perry* case provides an example here as well. Recall that the plaintiff was "sick too soon" for the eczema drug to have been the cause of his lymphoma. *Perry*, 564 F. Supp. 2d at 470.

## V. SETTING THE STANDARD FOR ADMISSIBILITY IS POLICY AS WELL AS SCIENCE

The Bradford Hill and validity indicia can help courts understand and articulate the grounds on which they admit or exclude expert evidence. It is important to emphasize, however, that they cannot establish a fixed standard for admissibility. The standard is, or at least should be, grounded in science. Under the Federal Rules of Evidence, *ipse dixit* should not do.[244] But how much evidence is needed is also a question of policy. Policy considerations most clearly come to the forefront in cases with little or no human data. Perhaps, therefore, it is not surprising that this set of cases sometimes produce conflicting specific-causation admissibility rulings.[245] And it is this set of cases where some courts seem to be most likely to admit expert-specific-causation testimony with limited scientific support.[246]

*In re Testosterone Replacement Therapy Products Liability Litigation Coordinated Pretrial Proceedings*[247] is a typical example. There, the court made the following statement:

> Specifically, AbbVie argues that plaintiffs' experts were required to quantify the risks posed by each alternative cause that they could not definitively rule out. The Court disagrees that such quantification is required to reliably conduct a reliable differential etiology. It is true that where numerous causes of an injury are plausible, an expert may not "simply pick [ ] the cause that is most advantageous to [the plaintiff's] claim." *Viterbo v. Dow Chem. Co.*, 826 F.2d 420, 424 (5th Cir. 1987). Quantifying the risks posed by alternative causes is, of course, one way to establish that a particular risk was a likely cause of the plaintiff's injury. But under Seventh Circuit law, in a case like this, the expert's task is

---

244. *See* Faigman & Mnookin, *supra* note 37, at 628. Especially when the scientific record is thin, the fine line between legitimate inference from incomplete evidence and inappropriate speculation and conjecture becomes both absolutely critical and particularly challenging to navigate.

At least one academic, however, has argued for this *ipse dixit* approach by, for example, permitting clinicians to testify as to their clinical judgement on specific causation. *See* Barbara Pfeffer Billauer, *The Causal Conundrum: Examining the Medical-Legal Disconnect in Toxic Tort Cases from a Cultural Perspective or How the Law Swallowed the Epidemiologist and Grew Long Legs and a Tail*, 51 CREIGHTON L. REV. 319 (2018). For a federal-court opinion that seems to come close to adopting this position, see *In re* Roundup Prods. Liab. Litig., 358 F. Supp. 3d 956, 960 (N.D. Cal. 2019).

245. *Compare* Messick v. Novartis Pharms. Corp., 747 F.3d 1193 (9th Cir. 2014), Bee v. Novartis Pharms. Corp., 18 F. Supp. 3d 268, 307 (E.D.N.Y. 2014), *and* Monroe v. Novartis Pharms. Corp., 29 F. Supp. 3d 1115, 1122 (S.D. Ohio 2014) (all admitting testimony), *with* Garrison v. Novartis Pharms. Corp., 30 F. Supp. 3d 1325, 1339 (M.D. Ala. 2014) (excluding testimony).

246. *See* Wendell v. Johnson & Johnson, No. C 09-cv-04124, 2014 WL 2943572, at *3 (N.D. Cal. June 30, 2014), *rev'd sub nom.* Wendell v. GlaxoSmithKline LLC, 858 F.3d 1227 (9th Cir. 2017).

247. *In re* Testosterone Replacement Therapy Prods. Liab. Litig. Coordinated Pretrial Proc., No. 14 C 1748, 2017 WL 1833173, at *17 (N.D. Ill. May 8, 2017).

> to provide a reliable basis for concluding that the drug at issue was a "substantial factor" in the development of the plaintiff's injury and that the other potential causes are unlikely to have been the injury's sole cause. Quantifying the risks of alternative causes is not required.[248]

Because of the "group to individual" G2i problem[249] confronting all specific-causation analyses, courts are confronted with a dilemma whenever there are multiple causes of a plaintiff's injury and there is no compelling quantitative evidence permitting an expert to offer a more-likely-than-not opinion. This court, and other courts have been reluctant to require the impossible from plaintiffs and ask for the same level of proof of causation that they seek in most general-causation inquiries or in many specific-causation inquiries where the data are more complete. But they are also reluctant to adopt any of the proposals found in the literature that call upon courts to abolish the specific-causation requirement altogether, to adopt something other than a but-for test, or to give plaintiffs a percentage recovery when the plaintiff cannot meet their burden of proof.[250]

The alternative is to lessen the level of justification required from an expert before the expert is permitted to make a specific-causation claim.[251] It is clear that courts sometimes allow an expert to express a causation opinion with less justification than they require for general causation.[252] When they do so they implicitly or explicitly leaven scientific rigor with legal-process policy goals.[253]

---

248. *See also In re* Neurontin Mktg., Sales Pracs., & Prods. Liab. Litig., 612 F. Supp. 2d 116, 156 (D. Mass. 2009); Adeghe v. Janssen Pharms., Inc., 16 Civ. 2235, 2017 WL 3741310, at *4 (S.D.N.Y. Aug. 30, 2017); *Wendell*, 858 F.3d at 1234. *But see McMunn v. Babcock & Wilcox Power Generation Group, Inc.*, 869 F.3d 246 (2d Cir. 2017) for a case where the plaintiff unsuccessfully attempted to apply this approach in a radiation-exposure situation.

249. *See supra* note 2 (discussing the G2i problem).

250. Although this last alternative may seem particularly attractive and has some precedent in the loss-of-chance cases such as *Matsuyama v. Birnbaum*, 890 N.E.2d 819 (Mass. 2008), it is only even theoretically possible in cases where there is valid epidemiologic evidence sufficient to establish general causation. In all the cases where there is insufficient epidemiology, there is no plausible way to arrive at a percentage that reflects the probability that exposure to the defendant's product caused the plaintiff's injury.

251. For a discussion of the level of justification required of an expert when making a causal assertion, see Joseph Sanders, *Science, Law, and the Expert Witness*, 72(1) L. & CONTEMP. PROBS. 63 (2009).

252. *See, e.g.*, Messick v. Novartis Pharms. Corp., 747 F.3d 1193, 1199 (9th Cir. 2014).

253. *See* Nancy Gertner & Joseph Sanders, *Alternatives to Traditional Adversary Methods of Presenting Scientific Expertise in the Legal System*, 147 DAEDALUS 135 (2018). In this regard, the courts' lower admissibility thresholds are similar to the lower threshold courts have often employed in admitting state forensic evidence in criminal cases. David L. Faigman, *Evidentiary Incommensurability: A Preliminary Exploration of the Problem of Reasoning from General Scientific Data to Individualized Legal Decision Making*, 75 BROOK. L. REV. 1115 (2010); Joseph Sanders, *Applying* Daubert *Inconsistently? Proof of Individual Causation in Toxic Tort and Forensic Cases*, 75 BROOK. L. REV. 1367 (2010).

Certainly, there is room for disagreement about how much evidence is required before the court concludes that the expert has sufficiently justified her conclusion. However, in this situation the reasoning of the court is just as important as the result. What is unfortunate is that many courts—those that do admit testimony based on limited evidence that the plaintiff's injury was caused by the defendant's substance—follow a path of least resistance embodied in the above *Testosterone Replacement Therapy Products Liability Litigation* quote.[254] They simply say that some higher standard, e.g., quantifiable data, is not required but then fail to engage in a serious discussion of the available evidence. Other opinions fall back on legal nostrums such as the statement that differential diagnosis is a recognized methodology, and therefore, testimony based on this methodology should be admitted.[255] Or they rely on the bald statement that the plaintiff's expert is not required to exclude all alternative causes, even when there are quite plausible causes that are not ruled out. Some opinions are so truncated and conclusory that it is not even possible to assess the merits of the expert's arguments to exclude other causes.[256] If it accomplishes nothing else, we hope this Article will encourage courts to employ the analytical tools provided by the Bradford Hill indicia to write more complete and cogent opinions justifying their ruling.

A second policy issue presents itself when a competing cause is by way of a personal background risk, especially a risk over which the individual has no control. From a scientific perspective, all risks are equal and work for or against proof of specific causation regardless of whether they are caused by a responsible agent. From a policy perspective, courts may choose not to treat some risk factors as competing causes, perhaps because they are innate or other factors of which an individual has little or no control or are otherwise in the range of normal human variation and expression, which the defendant's product would be expected to safely accommodate. We noted this possibility in our discussion of the competing risk factors of estrogen use and obesity.[257]

## CONCLUSION

Assessing specific causation in toxic-tort cases is one of the most difficult tasks confronting judges in a post-*Daubert* world. As a number of us have noted in earlier articles, this is primarily because it presents a classic G2i problem, a search for the causes of an effect rather than the effects of a cause. Because most available science does not directly address this issue, experts are compelled to approach the

---

254. *See supra* note 248.

255. This type of reasoning is particularly likely to be found in *Frye* states (*see* Marsh v. Valyou, 977 So. 2d 543 (Fla. 2007); Walsh v. BASF Corp., 191 A.3d 838 (Pa. Super. Ct. 2018)), but occasionally appears in federal cases as well (*see* Johnson v. Mead Johnson & Co., 754 F.3d 557, 564 (8th Cir. 2014)).

256. The number of cases that could be cited for this proposition is quite large. Here are two recent examples. Note that they often are published only as slip opinions. *See, e.g.*, McWilliams v. Novartis AG, No. 2:17-CV-14302-Rosenberg/Maynard, 2018 WL 3364617 (S.D. Fla. July 9, 2018); Holder v. Interlake S.S. Co., 16-CV-343-wmc, 2018 WL 1725694, at *17 (W.D. Wis. Apr. 10, 2018); Childress v. Johnson & Johnson, No. 2:12-CV-01564, 2017 WL 6350504, at *2 (S.D.W. Va. Dec. 12, 2017).

257. *See supra* p. 57.

problem indirectly by attempting to eliminate possible competing causes of an injury.

In this Article, we have discussed the evidence available to answer causal questions in toxic-tort cases and reviewed how this evidence may be brought to bear to answer the specific-causation question within the structure of current tort-law rules. The Bradford Hill indicia have previously been useful to scientists and courts in addressing the question of general causation. In this Article we hope we have demonstrated that these same guidelines, supplemented by considerations of internal and external validity, constitute a set of analytical tools that can enhance a court's ability to arrive at defensible and coherent specific-causation rulings. Unfortunately, in the process we have demonstrated that a complete and well-informed differential etiology is still often beyond our reach.