Reference Manual on Scientific Evidence

Third Edition

Committee on the Development of the Third Edition of the
Reference Manual on Scientific Evidence

Committee on Science, Technology, and Law
Policy and Global Affairs

FEDERAL JUDICIAL CENTER

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

Summary Table of Contents

A detailed Table of Contents appears at the front of each chapter.

Reference Guide on Epidemiology

MICHAEL D. GREEN, D. MICHAL FREEDMAN, AND
LEON GORDIS

Michael D. Green, J.D., is Bess & Walter Williams Chair in Law, Wake Forest University School of Law, Winston-Salem, North Carolina.

D. Michal Freedman, J.D., Ph.D., M.P.H., is Epidemiologist, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland.

Leon Gordis, M.D., M.P.H., Dr.P.H., is Professor Emeritus of Epidemiology, Johns Hopkins Bloomberg School of Public Health, and Professor Emeritus of Pediatrics, Johns Hopkins School of Medicine, Baltimore, Maryland.

CONTENTS

I. Introduction

Epidemiology is the field of public health and medicine that studies the incidence, distribution, and etiology of disease in human populations. The purpose of epidemiology is to better understand disease causation and to prevent disease in groups of individuals. Epidemiology assumes that disease is not distributed randomly in a group of individuals and that identifiable subgroups, including those exposed to certain agents, are at increased risk of contracting particular diseases.[1]

Judges and juries are regularly presented with epidemiologic evidence as the basis of an expert's opinion on causation.[2] In the courtroom, epidemiologic research findings are offered to establish or dispute whether exposure to an agent[3]

---

1. Although epidemiologists may conduct studies of beneficial agents that prevent or cure disease or other medical conditions, this reference guide refers exclusively to outcomes as diseases, because they are the relevant outcomes in most judicial proceedings in which epidemiology is involved.

2. Epidemiologic studies have been well received by courts deciding cases involving toxic substances. *See, e.g.,* Siharath v. Sandoz Pharms. Corp., 131 F. Supp. 2d 1347, 1356 (N.D. Ga. 2001) ("The existence of relevant epidemiologic studies can be a significant factor in proving general causation in toxic tort cases. Indeed, epidemiologic studies provide 'the primary generally accepted methodology for demonstrating a causal relation between a chemical compound and a set of symptoms or disease.'" (quoting Conde v. Velsicol Chem. Corp., 804 F. Supp. 972, 1025–26 (S.D. Ohio 1992))), *aff'd,* 295 F.3d 1194 (11th Cir. 2002); Berry v. CSX Transp., Inc., 709 So. 2d 552, 569 (Fla. Dist. Ct. App. 1998). Well-conducted studies are uniformly admitted. 3 Modern Scientific Evidence: The Law and Science of Expert Testimony § 23.1, at 187 (David L. Faigman et al. eds., 2007–08) [hereinafter Modern Scientific Evidence]. Since *Daubert v. Merrell Dow Pharmaceuticals,* 509 U.S. 579 (1993), the predominant use of epidemiologic studies is in connection with motions to exclude the testimony of expert witnesses. Cases deciding such motions routinely address epidemiology and its implications for the admissibility of expert testimony on causation. Often it is not the investigator who conducted the study who is serving as an expert witness in a case in which the study bears on causation. *See, e.g.,* Kennedy v. Collagen Corp., 161 F.3d 1226 (9th Cir. 1998) (physician is permitted to testify about causation); DeLuca v. Merrell Dow Pharms., Inc., 911 F.2d 941, 953 (3d Cir. 1990) (a pediatric pharmacologist expert's credentials are sufficient pursuant to Fed. R. Evid. 702 to interpret epidemiologic studies and render an opinion based thereon); Medalen v. Tiger Drylac U.S.A., Inc., 269 F. Supp. 2d 1118, 1129 (D. Minn. 2003) (holding toxicologist could testify to general causation but not specific causation); Burton v. R.J. Reynolds Tobacco Co., 181 F. Supp. 2d 1256, 1267 (D. Kan. 2002) (a vascular surgeon was permitted to testify to general causation); Landrigan v. Celotex Corp., 605 A.2d 1079, 1088 (N.J. 1992) (an epidemiologist was permitted to testify to both general causation and specific causation); Trach v. Fellin, 817 A.2d 1102, 1117–18 (Pa. Super. Ct. 2003) (an expert who was a toxicologist and pathologist was permitted to testify to general and specific causation).

3. We use the term "agent" to refer to any substance external to the human body that potentially causes disease or other health effects. Thus, drugs, devices, chemicals, radiation, and minerals (e.g., asbestos) are all agents whose toxicity an epidemiologist might explore. A single agent or a number of independent agents may cause disease, or the combined presence of two or more agents may be necessary for the development of the disease. Epidemiologists also conduct studies of individual characteristics, such as blood pressure and diet, which might pose risks, but those studies are rarely of interest in judicial proceedings. Epidemiologists also may conduct studies of drugs and other pharmaceutical products to assess their efficacy and safety.

caused a harmful effect or disease.[4] Epidemiologic evidence identifies agents that are associated with an increased risk of disease in groups of individuals, quantifies the amount of excess disease that is associated with an agent, and provides a profile of the type of individual who is likely to contract a disease after being exposed to an agent. Epidemiology focuses on the question of general causation (i.e., is the agent capable of causing disease?) rather than that of specific causation (i.e., did it cause disease in a particular individual?).[5] For example, in the 1950s, Doll and Hill and others published articles about the increased risk of lung cancer in cigarette smokers. Doll and Hill's studies showed that smokers who smoked 10 to 20 cigarettes a day had a lung cancer mortality rate that was about 10 times higher than that for nonsmokers.[6] These studies identified an association between smoking cigarettes and death from lung cancer that contributed to the determination that smoking causes lung cancer.

However, it should be emphasized that *an association is not equivalent to causation*.[7] An association identified in an epidemiologic study may or may not be

4. *E.g.,* Bonner v. ISP Techs., Inc., 259 F.3d 924 (8th Cir. 2001) (a worker exposed to organic solvents allegedly suffered organic brain dysfunction); Burton v. R.J. Reynolds Tobacco Co., 181 F. Supp. 2d 1256 (D. Kan. 2002) (cigarette smoking was alleged to have caused peripheral vascular disease); *In re* Bextra & Celebrex Mktg. Sales Practices & Prod. Liab. Litig., 524 F. Supp. 2d 1166 (N.D. Cal. 2007) (multidistrict litigation over drugs for arthritic pain that caused heart disease); Ruff v. Ensign-Bickford Indus., Inc., 168 F. Supp. 2d 1271 (D. Utah 2001) (chemicals that escaped from an explosives manufacturing site allegedly caused non-Hodgkin's lymphoma in nearby residents); Castillo v. E.I. du Pont De Nemours & Co., 854 So. 2d 1264 (Fla. 2003) (a child born with a birth defect allegedly resulting from mother's exposure to a fungicide).

5. This terminology and the distinction between general causation and specific causation are widely recognized in court opinions. *See, e.g.,* Norris v. Baxter Healthcare Corp., 397 F.3d 878 (10th Cir. 2005); *In re* Hanford Nuclear Reservation Litig., 292 F.3d 1124, 1129 (9th Cir. 2002) ("'Generic causation' has typically been understood to mean the capacity of a toxic agent . . . to cause the illnesses complained of by plaintiffs. If such capacity is established, 'individual causation' answers whether that toxic agent actually caused a particular plaintiff's illness."); *In re* Rezulin Prods. Liab. Litig., 369 F. Supp. 2d 398, 402 (S.D.N.Y. 2005); Soldo v. Sandoz Pharms. Corp., 244 F. Supp. 2d 434, 524–25 (W.D. Pa. 2003); Burton v. R.J. Reynolds Tobacco Co., 181 F. Supp. 2d 1256, 1266–67 (D. Kan. 2002). For a discussion of specific causation, see *infra* Section VII.

6. Richard Doll & A. Bradford Hill, *Lung Cancer and Other Causes of Death in Relation to Smoking: A Second Report on the Mortality of British Doctors,* 2 Brit. Med. J. 1071 (1956).

7. *See* Soldo v. Sandoz Pharms. Corp., 244 F. Supp. 2d 434, 461 (W.D. Pa. 2003) (Hill criteria [see *infra* Section V] developed to assess whether an association is causal); Miller v. Pfizer, Inc., 196 F. Supp. 2d 1062, 1079–80 (D. Kan. 2002); Magistrini v. One Hour Martinizing Dry Cleaning, 180 F. Supp. 2d 584, 591 (D.N.J. 2002) ("[A]n association is not equivalent to causation." (quoting the second edition of this reference guide)); Zandi v. Wyeth a/k/a Wyeth, Inc., No. 27-CV-06-6744, 2007 WL 3224242, at *11 (D. Minn. Oct. 15, 2007).

Association is more fully discussed *infra* Section III. The term is used to describe the relationship between two events (e.g., exposure to a chemical agent and development of disease) that occur more frequently together than one would expect by chance. Association does not necessarily imply a causal effect. Causation is used to describe the association between two events when one event is a necessary link in a chain of events that results in the effect. Of course, alternative causal chains may exist that do not include the agent but that result in the same effect. For general treatment of causation in tort law

causal.[8] Assessing whether an association is causal requires an understanding of the strengths and weaknesses of the study's design and implementation, as well as a judgment about how the study findings fit with other scientific knowledge. It is important to emphasize that all studies have "flaws" in the sense of limitations that add uncertainty about the proper interpretation of the results.[9] Some flaws are inevitable given the limits of technology, resources, the ability and willingness of persons to participate in a study, and ethical constraints. In evaluating epidemiologic evidence, the key questions, then, are the extent to which a study's limitations compromise its findings and permit inferences about causation.

A final caveat is that employing the results of group-based studies of risk to make a causal determination for an individual plaintiff is beyond the limits of epidemiology. Nevertheless, a substantial body of legal precedent has developed that addresses the use of epidemiologic evidence to prove causation for an individual litigant through probabilistic means, and the law developed in these cases is discussed later in this reference guide.[10]

The following sections of this reference guide address a number of critical issues that arise in considering the admissibility of, and weight to be accorded to, epidemiologic research findings. Over the past several decades, courts frequently have confronted the use of epidemiologic studies as evidence and have recognized their utility in proving causation. As the Third Circuit observed in *DeLuca v. Merrell Dow Pharmaceuticals, Inc.*: "The reliability of expert testimony founded on reasoning from epidemiologic data is generally a fit subject for judicial notice; epidemiology is a well-established branch of science and medicine, and epidemiologic evidence has been accepted in numerous cases."[11] Indeed,

and that for factual causation to exist an agent must be a necessary link in a causal chain sufficient for the outcome, see Restatement (Third) of Torts: Liability for Physical Harm § 26 (2010). Epidemiologic methods cannot deductively prove causation; indeed, all empirically based science cannot affirmatively prove a causal relation. *See, e.g.,* Stephan F. Lanes, *The Logic of Causal Inference in Medicine,* in Causal Inference 59 (Kenneth J. Rothman ed., 1988). However, epidemiologic evidence can justify an inference that an agent causes a disease. See *infra* Section V.

8. *See infra* Section IV.

9. *See In re* Phenylpropanolamine (PPA) Prods. Liab. Litig., 289 F. Supp. 2d 1230, 1240 (W.D. Wash. 2003) (quoting this reference guide and criticizing defendant's "ex post facto dissection" of a study); *In re* Orthopedic Bone Screw Prods. Liab. Litig., MDL No. 1014, 1997 U.S. Dist. LEXIS 6441, at ★26–★27 (E.D. Pa. May 5, 1997) (holding that despite potential for several biases in a study that "may . . . render its conclusions inaccurate," the study was sufficiently reliable to be admissible); Joseph L. Gastwirth, *Reference Guide on Survey Research,* 36 Jurimetrics J. 181, 185 (1996) (review essay) ("One can always point to a potential flaw in a statistical analysis.").

10. *See infra* Section VII.

11. 911 F.2d 941, 954 (3d Cir. 1990); *see also* Norris v. Baxter Healthcare Corp., 397 F.3d 878, 882 (10th Cir. 2005) (an extensive body of exonerative epidemiologic evidence must be confronted and the plaintiff must provide scientifically reliable contrary evidence); *In re* Meridia Prods. Liab. Litig., 328 F. Supp. 2d 791, 800 (N.D. Ohio 2004) ("Epidemiologic studies are the primary generally accepted methodology for demonstrating a causal relation between the chemical compound and a set of symptoms or a disease. . . ." (quoting Conde v. Velsicol Chem. Corp., 804 F. Supp. 972,

much more difficult problems arise for courts when there is a paucity of epidemiologic evidence.[12]

Three basic issues arise when epidemiology is used in legal disputes, and the methodological soundness of a study and its implications for resolution of the question of causation must be assessed:

1. Do the results of an epidemiologic study or studies reveal an association between an agent and disease?
2. Could this association have resulted from limitations of the study (bias, confounding, or sampling error), and, if so, from which?
3. Based on the analysis of limitations in Item 2, above, and on other evidence, how plausible is a causal interpretation of the association?

Section II explains the different kinds of epidemiologic studies, and Section III addresses the meaning of their outcomes. Section IV examines concerns about the methodological validity of a study, including the problem of sampling error.[13] Section V discusses general causation, considering whether an agent is capable of causing disease. Section VI deals with methods for combining the results of multiple epidemiologic studies and the difficulties entailed in extracting a single global measure of risk from multiple studies. Additional legal questions that arise in most toxic substances cases are whether population-based epidemiologic evidence can be used to infer specific causation, and, if so, how. Section VII addresses specific causation—the matter of whether a specific agent caused the disease in a given plaintiff.

1025–26 (S.D. Ohio 1992))); Brasher v. Sandoz Pharms. Corp., 160 F. Supp. 2d 1291, 1296 (N.D. Ala. 2001) ("Unquestionably, epidemiologic studies provide the best proof of the general association of a particular substance with particular effects, but it is not the only scientific basis on which those effects can be predicted.").

12. *See infra* note 181.

13. For a more in-depth discussion of the statistical basis of epidemiology, see David H. Kaye & David A. Freedman, Reference Guide on Statistics, Section II.A, in this manual, and two case studies: Joseph Sanders, *The Bendectin Litigation: A Case Study in the Life Cycle of Mass Torts,* 43 Hastings L.J. 301 (1992); Devra L. Davis et al., *Assessing the Power and Quality of Epidemiologic Studies of Asbestos-Exposed Populations,* 1 Toxicological & Indus. Health 93 (1985). *See also* References on Epidemiology and References on Law and Epidemiology at the end of this reference guide.

II. What Different Kinds of Epidemiologic Studies Exist?

A. Experimental and Observational Studies of Suspected Toxic Agents

To determine whether an agent is related to the risk of developing a certain disease or an adverse health outcome, we might ideally want to conduct an experimental study in which the subjects would be randomly assigned to one of two groups: one group exposed to the agent of interest and the other not exposed. After a period of time, the study participants in both groups would be evaluated for the development of the disease. This type of study, called a randomized trial, clinical trial, or true experiment, is considered the gold standard for determining the relationship of an agent to a health outcome or adverse side effect. Such a study design is often used to evaluate new drugs or medical treatments and is the best way to ensure that any observed difference in outcome between the two groups is likely to be the result of exposure to the drug or medical treatment.

Randomization minimizes the likelihood that there are differences in relevant characteristics between those exposed to the agent and those not exposed. Researchers conducting clinical trials attempt to use study designs that are placebo controlled, which means that the group not receiving the active agent or treatment is given an inactive ingredient that appears similar to the active agent under study. They also use double blinding where possible, which means that neither the participants nor those conducting the study know which group is receiving the agent or treatment and which group is given the placebo. However, ethical and practical constraints limit the use of such experimental methodologies to assess the value of agents that are thought to be beneficial to human beings.[14]

When an agent's effects are suspected to be harmful, researchers cannot knowingly expose people to the agent.[15] Instead epidemiologic studies typically

---

14. Although experimental human studies cannot intentionally expose subjects to toxins, they can provide evidence that a new drug or other beneficial intervention also has adverse effects. *See In re* Bextra & Celebrex Mktg. Sales Practices & Prod. Liab. Litig., 524 F. Supp. 2d 1166, 1181 (N.D. Cal. 2007) (the court relied on a clinical study of Celebrex that revealed increased cardiovascular risk to conclude that the plaintiff's experts' testimony on causation was admissible); McDarby v. Merck & Co., 949 A.2d 223 (N.J. Super. Ct. App. Div. 2008) (explaining how clinical trials of Vioxx revealed an association with heart disease).

15. Experimental studies in which human beings are exposed to agents known or thought to be toxic are ethically proscribed. *See* Glastetter v. Novartis Pharms. Corp., 252 F.3d 986, 992 (8th Cir. 2001); Brasher v. Sandoz Pharms. Corp., 160 F. Supp. 2d 1291, 1297 (N.D. Ala. 2001). Experimental studies can be used where the agent under investigation is believed to be beneficial, as is the case in the development and testing of new pharmaceutical drugs. *See, e.g.,* McDarby v. Merck & Co., 949 A.2d 223, 270 (N.J. Super. Ct. App. Div. 2008) (an expert witness relied on a clinical trial of a new drug to find the adjusted risk for the plaintiff); *see also* Gordon H. Guyatt, *Using Randomized Trials in*

"observe"[16] a group of individuals who have been exposed to an agent of interest, such as cigarette smoke or an industrial chemical and compare them with another group of individuals who have not been exposed. Thus, the investigator identifies a group of subjects who have been exposed[17] and compares their rate of disease or death with that of an unexposed group. In contrast to clinical studies in which potential risk factors can be controlled, epidemiologic investigations generally focus on individuals living in the community, for whom characteristics other than the one of interest, such as diet, exercise, exposure to other environmental agents, and genetic background, may distort a study's results. Because these characteristics cannot be controlled directly by the investigator, the investigator addresses their possible role in the relationship being studied by considering them in the design of the study and in the analysis and interpretation of the study results (see *infra* Section IV).[18] We emphasize that the Achilles' heel of observational studies is the possibility of differences in the two populations being studied with regard to risk factors other than exposure to the agent.[19] By contrast, experimental studies, in which subjects are randomized, generally avoid this problem.

B. Types of Observational Study Design

Several different types of observational epidemiologic studies can be conducted.[20] Study designs may be chosen because of suitability for investigating the question of interest, timing constraints, resource limitations, or other considerations.

Most observational studies collect data about both exposure and health outcome in every individual in the study. The two main types of observational studies are cohort studies and case-control studies. A third type of observational study is a cross-sectional study, although cross-sectional studies are rarely useful in identifying toxic agents.[21] A final type of observational study, one in which data about

---

Pharmacoepidemiology, in Drug Epidemiology and Post-Marketing Surveillance 59 (Brian L. Strom & Giampaolo Velo eds., 1992). Experimental studies also may be conducted that entail the discontinuation of exposure to a harmful agent, such as studies in which smokers are randomly assigned to a variety of smoking cessation programs or have no cessation.

16. Classifying these studies as observational in contrast to randomized trials can be misleading to those who are unfamiliar with the area, because subjects in a randomized trial are observed as well. Nevertheless, the use of the term "observational studies" to distinguish them from experimental studies is widely employed.

17. The subjects may have voluntarily exposed themselves to the agent of interest, as is the case, for example, for those who smoke cigarettes, or subjects may have been exposed involuntarily or even without knowledge to an agent, such as in the case of employees who are exposed to chemical fumes at work.

18. *See* David A. Freedman, *Oasis or Mirage?* 21 Chance 59, 59–61 (Mar. 2008).

19. Both experimental and observational studies are subject to random error. See *infra* Section IV.A.

20. Other epidemiologic studies collect data about the group as a whole, rather than about each individual in the group. These group studies are discussed *infra* Section II.B.4.

21. See *infra* Section II.B.3.

individuals are not gathered, but rather population data about exposure and disease are used, is an ecological study.[22]

The difference between cohort studies and case-control studies is that cohort studies measure and compare the incidence of disease in the exposed and unexposed ("control") groups, while case-control studies measure and compare the frequency of exposure in the group with the disease (the "cases") and the group without the disease (the "controls"). In a case-control study, the rates of exposure in the cases and the rates in the controls are compared, and the odds of having the disease when exposed to a suspected agent can be compared with the odds when not exposed. The critical difference between cohort studies and case-control studies is that cohort studies begin with exposed people and unexposed people, while case-control studies begin with individuals who are selected based on whether they have the disease or do not have the disease and their exposure to the agent in question is measured. The goal of both types of studies is to determine if there is an association between exposure to an agent and a disease and the strength (magnitude) of that association.

1. Cohort studies

In cohort studies,[23] researchers define a study population without regard to the participants' disease status. The cohort may be defined in the present and followed forward into the future (prospectively) or it may be constructed retrospectively as of sometime in the past and followed over historical time toward the present. In either case, the researchers classify the study participants into groups based on whether they were exposed to the agent of interest (see Figure 1).[24] In a prospective study, the exposed and unexposed groups are followed for a specified length of time, and the proportions of individuals in each group who develop the disease of interest are compared. In a retrospective study, the researcher will determine the proportion of individuals in the exposed group who developed the disease from available records or evidence and compare that proportion with the proportion of another group that was not exposed.[25] Thus, as illustrated in Table 1,

---

22. For thumbnail sketches on all types of epidemiologic study designs, see Brian L. Strom, *Study Designs Available for Pharmacoepidemiology Studies, in* Pharmacoepidemiology 17, 21–26 (Brian L. Strom ed., 4th ed. 2005).

23. Cohort studies also are referred to as prospective studies and followup studies.

24. In some studies, there may be several groups, each with a different magnitude of exposure to the agent being studied. Thus, a study of cigarette smokers might include heavy smokers (>3 packs a day), moderate smokers (1 to 2 packs a day), and light smokers (<1 pack a day). *See, e.g.,* Robert A. Rinsky et al., *Benzene and Leukemia: An Epidemiologic Risk Assessment,* 316 New Eng. J. Med. 1044 (1987).

25. Sometimes in retrospective cohort studies the researcher gathers historical data about exposure and disease outcome of a cohort. Harold A. Kahn, An Introduction to Epidemiologic Methods 39–41 (1983). Irving Selikoff, in his seminal study of asbestotic disease in insulation workers, included several hundred workers who had died before he began the study. Selikoff was able to obtain information about exposure from union records and information about disease from hospital and autopsy

Figure 1. Design of a cohort study.
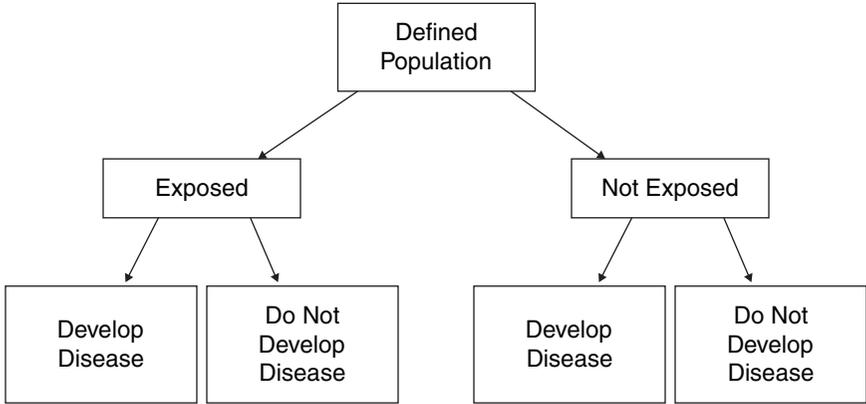
Table 1. Cross-Tabulation of Exposure by Disease Status

	No Disease	Disease	Totals	Incidence Rates of Disease
Not exposed	a	c	$a + c$	$c/(a + c)$
Exposed	b	d	$b + d$	$d/(b + d)$

a researcher would compare the proportion of unexposed individuals with the disease, $c/(a + c)$, with the proportion of exposed individuals with the disease, $d/(b + d)$. If the exposure causes the disease, the researcher would expect a greater proportion of the exposed individuals to develop the disease than the unexposed individuals.[26]

One advantage of the cohort study design is that the temporal relationship between exposure and disease can often be established more readily than in other study designs, especially a case-control design, discussed below. By tracking people who are initially not affected by the disease, the researcher can determine the time of disease onset and its relation to exposure. This temporal relationship is critical to the question of causation, because exposure must precede disease onset if exposure caused the disease.

As an example, in 1950 a cohort study was begun to determine whether uranium miners exposed to radon were at increased risk for lung cancer as com-

records. Irving J. Selikoff et al., *The Occurrence of Asbestosis Among Insulation Workers in the United States*, 132 Ann. N.Y. Acad. Sci. 139, 143 (1965).

    26. Researchers often examine the rate of disease or death in the exposed and control groups. The rate of disease or death entails consideration of the number developing disease within a specified period. All smokers and nonsmokers will, if followed for 100 years, die. Smokers will die at a greater rate than nonsmokers in the earlier years.

pared with nonminers. The study group (also referred to as the exposed cohort) consisted of 3400 white, underground miners. The control group (which need not be the same size as the exposed cohort) comprised white nonminers from the same geographic area. Members of the exposed cohort were examined every 3 years, and the degree of this cohort's exposure to radon was measured from samples taken in the mines. Ongoing testing for radioactivity and periodic medical monitoring of lungs permitted the researchers to examine whether disease was linked to prior work exposure to radiation and allowed them to discern the relationship between exposure to radiation and disease. Exposure to radiation was associated with the development of lung cancer in uranium miners.[27]

The cohort design is used often in occupational studies such as the one just discussed. Because the design is not experimental, and the investigator has no control over what other exposures a subject in the study may have had, an increased risk of disease among the exposed group may be caused by agents other than the exposure of interest. A cohort study of workers in a certain industry that pays below-average wages might find a higher risk of cancer in those workers. This may be because they work in that industry, or, among other reasons, because low-wage groups are exposed to other harmful agents, such as environmental toxins present in higher concentrations in their neighborhoods. In the study design, the researcher must attempt to identify factors other than the exposure that may be responsible for the increased risk of disease. If data are gathered on other possible etiologic factors, the researcher generally uses statistical methods[28] to assess whether a true association exists between working in the industry and cancer. Evaluating whether the association is causal involves additional analysis, as discussed in Section V.

2. Case-control studies

In case-control studies,[29] the researcher begins with a group of individuals who have a disease (cases) and then selects a similar group of individuals who do not have the disease (controls). (Ideally, controls should come from the same source population as the cases.) The researcher then compares the groups in terms of past exposures. If a certain exposure is associated with or caused the disease, a higher proportion of past exposure among the cases than among the controls would be expected (see Figure 2).

---

27. This example is based on a study description in Abraham M. Lilienfeld & David E. Lilienfeld, Foundations of Epidemiology 237–39 (2d ed. 1980). The original study is Joseph K. Wagoner et al., *Radiation as the Cause of Lung Cancer Among Uranium Miners,* 273 New Eng. J. Med. 181 (1965).

28. *See* Daniel L. Rubinfeld, Reference Guide on Multiple Regression, Section II.B, in this manual; David H. Kaye & David A. Freedman, Reference Guide on Statistics, Section V.D, in this manual.

29. Case-control studies are also referred to as retrospective studies, because researchers gather historical information about rates of exposure to an agent in the case and control groups.

Figure 2. Design of a case-control study.

Thus, for example, in the late 1960s, doctors in Boston were confronted with an unusual number of young female patients with vaginal adenocarcinoma. Those patients became the "cases" in a case-control study (because they had the disease in question) and were matched with "controls," who did not have the disease. Controls were selected based on their being born in the same hospitals and at the same time as the cases. The cases and controls were compared for exposure to agents that might be responsible, and researchers found maternal ingestion of DES (diethylstilbestrol) in all but one of the cases but none of the controls.[30]

An advantage of the case-control study is that it usually can be completed in less time and with less expense than a cohort study. Case-control studies are also particularly useful in the study of rare diseases, because if a cohort study were conducted, an extremely large group would have to be studied in order to observe the development of a sufficient number of cases for analysis.[31] A number of potential problems with case-control studies are discussed in Section IV.B.

3. Cross-sectional studies

A third type of observational study is a cross-sectional study. In this type of study, individuals are interviewed or examined, and the presence of both the exposure of interest and the disease of interest is determined in each individual at a single point in time. Cross-sectional studies determine the presence (prevalence) of both exposure and disease in the subjects and do not determine the development of disease or risk of disease (incidence). Moreover, because both exposure and disease are determined in an individual at the same point in time, it is not possible to establish the temporal relation between exposure and disease—that is, that the

---

30. *See* Arthur L. Herbst et al., *Adenocarcinoma of the Vagina: Association of Maternal Stilbestrol Therapy with Tumor Appearance,* 284 New Eng. J. Med. 878 (1971).

31. Thus, for example, to detect a doubling of disease caused by exposure to an agent where the incidence of disease is 1 in 100 in the unexposed population would require sample sizes of 3100 for the exposed and nonexposed groups for a cohort study, but only 177 for the case and control groups in a case-control study. Harold A. Kahn & Christopher T. Sempos, Statistical Methods in Epidemiology 66 (1989).

exposure preceded the disease, which would be necessary for drawing any causal inference. Thus, a researcher may use a cross-sectional study to determine the connection between a personal characteristic that does not change over time, such as blood type, and existence of a disease, such as aplastic anemia, by examining individuals and determining their blood types and whether they suffer from aplastic anemia. Cross-sectional studies are infrequently used when the exposure of interest is an environmental toxic agent (current smoking status is a poor measure of an individual's history of smoking), but these studies can provide valuable leads to further directions for research.[32]

4. Ecological studies

Up to now, we have discussed studies in which data on both exposure and health outcome are obtained for each individual included in the study.[33] In contrast, studies that collect data only about the group as a whole are called ecological studies.[34] In ecological studies, information about individuals is generally not gathered; instead, overall rates of disease or death for different groups are obtained and compared. The objective is to identify some difference between the two groups, such as diet, genetic makeup, or alcohol consumption, that might explain differences in the risk of disease observed in the two groups.[35] Such studies may be useful for identifying associations, but they rarely provide definitive causal answers.[36] The difficulty is illustrated below with an ecological study of the relationship between dietary fat and cancer.

---

32. For more information (and references) about cross-sectional studies, see Leon Gordis, Epidemiology 195–98 (4th ed. 2009).

33. Some individual studies may be conducted in which all members of a group or community are treated as exposed to an agent of interest (e.g., a contaminated water system) and disease status is determined individually. These studies should be distinguished from ecological studies.

34. In *Cook v. Rockwell International Corp.,* 580 F. Supp. 2d 1071, 1095–96 (D. Colo. 2006), the plaintiffs' expert conducted an ecological study in which he compared the incidence of two cancers among those living in a specified area adjacent to the Rocky Flats Nuclear Weapons Plant with other areas more distant. (The likely explanation for relying on this type of study is the time and expense of a study that gathered information about each individual in the affected area.) The court recognized that ecological studies are less probative than studies in which data are based on individuals but nevertheless held that limitation went to the weight of the study. Plaintiff's expert was permitted to testify to causation, relying on the ecological study he performed.

In *Renaud v. Martin Marietta Corp.,* 749 F. Supp. 1545, 1551 (D. Colo. 1990), *aff'd,* 972 F.2d 304 (10th Cir. 1992), the plaintiffs attempted to rely on an excess incidence of cancers in their neighborhood to prove causation. Unfortunately, the court confused the role of epidemiology in proving causation with the issue of the plaintiffs' exposure to the alleged carcinogen and never addressed the evidentiary value of the plaintiffs' evidence of a disease cluster (i.e., an unusually high incidence of a particular disease in a neighborhood or community). *Id.* at 1554.

35. David E. Lilienfeld & Paul D. Stolley, Foundations of Epidemiology 12 (3d ed. 1994).

36. Thus, the emergence of a cluster of adverse events associated with use of heparin, a longtime and widely-prescribed anticoagulent, led to suspicions that some specific lot of heparin was responsible. These concerns led the Centers for Disease Control to conduct a case control study that concluded

If a researcher were interested in determining whether a high dietary fat intake is associated with breast cancer, he or she could compare different countries in terms of their average fat intakes and their average rates of breast cancer. If a country with a high average fat intake also tends to have a high rate of breast cancer, the finding would suggest an association between dietary fat and breast cancer. However, such a finding would be far from conclusive, because it lacks particularized information about an individual's exposure and disease status (i.e., whether an individual with high fat intake is more likely to have breast cancer).[37] In addition to the lack of information about an individual's intake of fat, the researcher does not know about the individual's exposures to other agents (or other factors, such as a mother's age at first birth) that may also be responsible for the increased risk of breast cancer. This lack of information about each individual's exposure to an agent and disease status detracts from the usefulness of the study and can lead to an erroneous inference about the relationship between fat intake and breast cancer, a problem known as an ecological fallacy. The fallacy is assuming that, on average, the individuals in the study who have suffered from breast cancer consumed more dietary fat than those who have not suffered from the disease. This assumption may not be true. Nevertheless, the study is useful in that it identifies an area for further research: the fat intake of individuals who have breast cancer as compared with the fat intake of those who do not. Researchers who identify a difference in disease or death in an ecological study may follow up with a study based on gathering data about individuals.

Another epidemiologic approach is to compare disease rates over time and focus on disease rates before and after a point in time when some event of interest took place.[38] For example, thalidomide's teratogenicity (capacity to cause birth defects) was discovered after Dr. Widukind Lenz found a dramatic increase in the incidence of limb reduction birth defects in Germany beginning in 1960. Yet, other than with such powerful agents as thalidomide, which increased the incidence of limb reduction defects by several orders of magnitude, these secular-trend studies (also known as time-line studies) are less reliable and less able to

---

that contaminated heparin manufactured by Baxter was responsible for the outbreak of adverse events. See David B. Blossom et al., *Outbreak of Adverse Event Reactions Associated with Contaminated Heparin,* 359 New Eng. J. Med. 2674 (2008); *In re* Heparin Prods. Liab. Litig. 2011 WL 2971918 (N.D. Ohio July 21, 2011).

37. For a discussion of the data on this question and what they might mean, see David Freedman et al., Statistics (4th ed. 2007).

38. In *Wilson v. Merrell Dow Pharmaceuticals, Inc.,* 893 F.2d 1149, 1152–53 (10th Cir. 1990), the defendant introduced evidence showing total sales of Bendectin and the incidence of birth defects during the 1970–1984 period. In 1983, Bendectin was removed from the market, but the rate of birth defects did not change. The Tenth Circuit affirmed the lower court's ruling that the time-line data were admissible and that the defendant's expert witnesses could rely on them in rendering their opinions. Similar evidence was relied on in cases involving cell phones and the drug Parlodel, which was alleged to cause postpartum strokes in women who took the drug to suppress lactation. *See* Newman v. Motorola, Inc., 218 F. Supp. 2d 769, 778 (D. Md. 2002); Siharath v. Sandoz Pharms. Corp., 131 F. Supp. 2d 1347, 1358 (N.D. Ga. 2001).

detect modest causal effects than the observational studies described above. Other factors that affect the measurement or existence of the disease, such as improved diagnostic techniques and changes in lifestyle or age demographics, may change over time. If those factors can be identified and measured, it may be possible to control for them with statistical methods. Of course, unknown factors cannot be controlled for in these or any other kind of epidemiologic studies.

C. Epidemiologic and Toxicologic Studies

In addition to observational epidemiology, toxicology models based on live animal studies (in vivo) may be used to determine toxicity in humans.[39] Animal studies have a number of advantages. They can be conducted as true experiments, and researchers control all aspects of the animals' lives. Thus, they can avoid the problem of confounding,[40] which epidemiology often confronts. Exposure can be carefully controlled and measured. Refusals to participate in a study are not an issue, and loss to followup very often is minimal. Ethical limitations are diminished, and animals can be sacrificed and their tissues examined, which may improve the accuracy of disease assessment. Animal studies often provide useful information about pathological mechanisms and play a complementary role to epidemiology by assisting researchers in framing hypotheses and in developing study designs for epidemiologic studies.

Animal studies have two significant disadvantages, however. First, animal study results must be extrapolated to another species—human beings—and differences in absorption, metabolism, and other factors may result in interspecies variation in responses. For example, one powerful human teratogen, thalidomide, does not cause birth defects in most rodent species.[41] Similarly, some known teratogens in animals are not believed to be human teratogens. In general, it is often difficult to confirm that an agent known to be toxic in animals is safe for human beings.[42] The second difficulty with inferring human causation from animal studies is that the high doses customarily used in animal studies require consideration of the dose–response relationship and whether a threshold no-effect dose exists.[43] Those matters are almost always fraught with considerable, and currently unresolvable, uncertainty.[44]

---

39. For an in-depth discussion of toxicology, see Bernard D. Goldstein & Mary Sue Henifin, Reference Guide on Toxicology, in this manual.

40. *See infra* Section IV.C.

41. Phillip Knightley et al., Suffer the Children: The Story of Thalidomide 271–72 (1979).

42. *See* Ian C.T. Nesbit & Nathan J. Karch, Chemical Hazards to Human Reproduction 98–106 (1983); Int'l Agency for Research on Cancer (IARC), Interpretation of Negative Epidemiologic Evidence for Carcinogenicity (N.J. Wald & Richard Doll eds., 1985) [hereafter IARC].

43. *See infra* Section V.C & note 119.

44. *See* Soldo v. Sandoz Pharms. Corp., 244 F. Supp. 2d 434, 466 (W.D. Pa. 2003) (quoting this reference guide in the first edition of the Reference Manual); *see also* General Elec. Co. v. Joiner, 522 U.S. 136, 143–45 (1997) (holding that the district court did not abuse its discretion in exclud-

Toxicologists also use in vitro methods, in which human or animal tissue or cells are grown in laboratories and are exposed to certain substances. The problem with this approach is also extrapolation—whether one can generalize the findings from the artificial setting of tissues in laboratories to whole human beings.[45]

Often toxicologic studies are the only or best available evidence of toxicity.[46] Epidemiologic studies are difficult, time-consuming, expensive, and sometimes, because of limited exposure or the infrequency of disease, virtually impossible to perform.[47] Consequently, they do not exist for a large array of environmental agents. Where both animal toxicologic and epidemiologic studies are available, no universal rules exist for how to interpret or reconcile them.[48] Careful assess-

---

ing expert testimony on causation based on expert's failure to explain how animal studies supported expert's opinion that agent caused disease in humans).

45. For a further discussion of these issues, see Bernard D. Goldstein & Mary Sue Henifin, Reference Guide on Toxicology, Section III.A, in this manual.

46. IARC, a well-regarded international public health agency, evaluates the human carcino-genicity of various agents. In doing so, IARC obtains all of the relevant evidence, including animal studies as well as any human studies. On the basis of a synthesis and evaluation of that evidence, IARC publishes a monograph containing that evidence and its analysis of the evidence and pro-vides a categorical assessment of the likelihood the agent is carcinogenic. In a preamble to each of its monographs, IARC explains what each of the categorical assessments means. Solely on the basis of the strength of animal studies, IARC may classify a substance as "probably carcinogenic to humans." International Agency for Research on Cancer, *Human Papillomaviruses,* 90 Monographs on the Evaluation of Carcinogenic Risks to Humans 9–10 (2007), *available at* http://monographs.iarc.fr/ ENG/Monographs/vol90/index.php; *see also* Magistrini v. One Hour Martinizing Dry Cleaning, 180 F. Supp. 2d 584, 600 n.18 (D.N.J. 2002). When IARC monographs are available, they are gener-ally recognized as authoritative. Unfortunately, IARC has conducted evaluations of only a fraction of potentially carcinogenic agents, and many suspected toxic agents cause effects other than cancer.

47. Thus, in a series of cases involving Parlodel, a lactation suppressant for mothers of newborns, efforts to conduct an epidemiologic study of its effect on causing strokes were stymied by the infre-quency of such strokes in women of child-bearing age. *See, e.g.,* Brasher v. Sandoz Pharms. Corp., 160 F. Supp. 2d 1291, 1297 (N.D. Ala. 2001). In other cases, a plaintiff's exposure to an overdose of a drug may be unique or nearly so. *See* Zuchowicz v. United States, 140 F.3d 381 (2d Cir. 1998).

48. *See* IARC, *supra* note 41 (identifying a number of substances and comparing animal toxicol-ogy evidence with epidemiologic evidence); Michele Carbone et al., *Modern Criteria to Establish Human Cancer Etiology,* 64 Cancer Res. 5518, 5522 (2004) (National Cancer Institute symposium concluding that "There should be no hierarchy [among different types of scientific methods to determine cancer causation]. Epidemiology, animal, tissue culture and molecular pathology should be seen as integrating evidences in the determination of human carcinogenicity.")

A number of courts have grappled with the role of animal studies in proving causation in a toxic substance case. One line of cases takes a very dim view of their probative value. For example, in *Brock v. Merrell Dow Pharmaceuticals, Inc.,* 874 F.2d 307, 313 (5th Cir. 1989), the court noted the "very limited usefulness of animal studies when confronted with questions of toxicity." A similar view is reflected in *Richardson v. Richardson-Merrell, Inc.,* 857 F.2d 823, 830 (D.C. Cir. 1988), Bell v. Swift Adhesives, Inc., 804 F. Supp. 1577, 1579–80 (S.D. Ga. 1992), and *Cadarian v. Merrell Dow Pharmaceuticals, Inc.,* 745 F. Supp. 409, 412 (E.D. Mich. 1989).

Other courts have been more amenable to the use of animal toxicology in proving causation. Thus, in *Marder v. G.D. Searle & Co.,* 630 F. Supp. 1087, 1094 (D. Md. 1986), *aff'd sub nom. Wheelahan v. G.D. Searle & Co.,* 814 F.2d 655 (4th Cir. 1987), the court observed: "There is a range of scientific

ment of the methodological validity and power[49] of the epidemiologic evidence must be undertaken, and the quality of the toxicologic studies and the questions of interspecies extrapolation and dose–response relationship must be considered.[50]

---

methods for investigating questions of causation—for example, toxicology and animal studies, clinical research, and epidemiology—which all have distinct advantages and disadvantages." In *Milward v. Acuity Specialty Products Group, Inc.,* 639 F.3d 11, 17–19 (1st Cir. 2011), the court endorsed an expert's use of a "weight-of-the-evidence" methodology, holding that the district court abused its discretion in ruling inadmissible an expert's testimony about causation based on that methodology. As a corollary to recognizing weight of the evidence as a valid scientific technique, the court also noted the role of judgment in making an appropriate inference from the evidence. While recognizing the legitimacy of the methodology, the court also acknowledged that, as with any scientific technique, it can be improperly applied. *See also* Metabolife Int'l, Inc. v. Wornick, 264 F.3d 832, 842 (9th Cir. 2001) (holding that the lower court erred in per se dismissing animal studies, which must be examined to determine whether they are appropriate as a basis for causation determination); *In re* Heparin Prods. Liab. Litig. 2011 WL 2971918 (N.D. Ohio July 21, 2011) (holding that animal toxicology in conjunction with other non-epidemiologic evidence can be sufficient to prove causation); Ruff v. Ensign-Bickford Indus., Inc., 168 F. Supp. 2d 1271, 1281 (D. Utah 2001) (affirming animal studies as sufficient basis for opinion on general causation.); *cf. In re* Paoli R.R. Yard PCB Litig., 916 F.2d 829, 853–54 (3d Cir. 1990) (questioning the exclusion of animal studies by the lower court). The Third Circuit in a subsequent opinion in *Paoli* observed:

> [I]n order for animal studies to be admissible to prove causation in humans, there must be good grounds to extrapolate from animals to humans, just as the methodology of the studies must constitute good grounds to reach conclusions about the animals themselves. Thus, the requirement of reliability, or "good grounds," extends to each step in an expert's analysis all the way through the step that connects the work of the expert to the particular case.

In re Paoli R.R. Yard PCB Litig., 35 F.3d 717, 743 (3d Cir. 1994); *see also* Cavallo v. Star Enter., 892 F. Supp. 756, 761–63 (E.D. Va. 1995) (courts must examine each of the steps that lead to an expert's opinion), *aff'd in part and rev'd in part,* 100 F.3d 1150 (4th Cir. 1996).

One explanation for these conflicting lines of cases may be that when there is a substantial body of epidemiologic evidence that addresses the causal issue, animal toxicology has much less probative value. That was the case, for example, in the Bendectin cases of *Richardson, Brock,* and *Cadarian.* Where epidemiologic evidence is not available, animal toxicology may be thought to play a more prominent role in resolving a causal dispute. *See* Michael D. Green, *Expert Witnesses and Sufficiency of Evidence in Toxic Substances Litigation: The Legacy of Agent Orange and Bendectin Litigation,* 86 Nw. U. L. Rev. 643, 680–82 (1992) (arguing that plaintiffs should be required to prove causation by a preponderance of the available evidence); Turpin v. Merrell Dow Pharms., Inc., 959 F.2d 1349, 1359 (6th Cir. 1992); *In re* Paoli R.R. Yard PCB Litig., No. 86-2229, 1992 U.S. Dist. LEXIS 16287, at *16 (E.D. Pa. 1992). For another explanation of these cases, see Gerald W. Boston, *A Mass-Exposure Model of Toxic Causation: The Control of Scientific Proof and the Regulatory Experience,* 18 Colum. J. Envtl. L. 181 (1993) (arguing that epidemiologic evidence should be required in mass-exposure cases but not in isolated-exposure cases); *see also* IARC, *supra* note 41; Bernard D. Goldstein & Mary Sue Henifin, Reference Guide on Toxicology, Section I.F, in this manual. The Supreme Court, in *General Electric Co. v. Joiner,* 522 U.S. 136, 144–45 (1997), suggested that there is no categorical rule for toxicologic studies, observing, "[W]hether animal studies can ever be a proper foundation for an expert's opinion [is] not the issue. . . . The [animal] studies were so dissimilar to the facts presented in this litigation that it was not an abuse of discretion for the District Court to have rejected the experts' reliance on them."

49. *See infra* Section IV.A.3.
50. *See* Ellen F. Heineman & Shelia Hoar Zahm, *The Role of Epidemiology in Hazard Evaluation,* 9 Toxic Substances J. 255, 258–62 (1989).

III. How Should Results of an Epidemiologic Study Be Interpreted?

Epidemiologists are ultimately interested in whether a causal relationship exists between an agent and a disease. However, the first question an epidemiologist addresses is whether an association exists between exposure to the agent and disease. An association between exposure to an agent and disease exists when they occur together more frequently than one would expect by chance.[51] Although a causal relationship is one possible explanation for an observed association between an exposure and a disease, an association does not necessarily mean that there is a cause–effect relationship. Interpreting the meaning of an observed association is discussed below.

This section begins by describing the ways of expressing the existence and strength of an association between exposure and disease. It reviews ways in which an incorrect result can be produced because of the sampling methods used in all observational epidemiologic studies and then examines statistical methods for evaluating whether an association is real or the result of a sampling error.

The strength of an association between exposure and disease can be stated in various ways,[52] including as a relative risk, an odds ratio, or an attributable risk.[53] Each of these measurements of association examines the degree to which the risk of disease increases when individuals are exposed to an agent.

A. Relative Risk

A commonly used approach for expressing the association between an agent and disease is relative risk ("RR"). It is defined as the ratio of the incidence rate (often referred to as incidence) of disease in exposed individuals to the incidence rate in unexposed individuals:

$$RR = \frac{(\text{Incidence rate in the exposed})}{(\text{Incidence rate in the unexposed})}$$

---

51. A negative association implies that the agent has a protective or curative effect. Because the concern in toxic substances litigation is whether an agent caused disease, this reference guide focuses on positive associations.

52. Another outcome measure is a risk difference. A risk difference is the difference between the proportion of disease in those exposed to the agent and the proportion of disease in those who were unexposed. Thus, in the example of relative risk in the text below discussing relative risk, the proportion of disease in those exposed is 40/100 and the proportion of disease in the unexposed is 20/100. The risk difference is 20/100.

53. Numerous courts have employed these measures of the strength of an association. *See, e.g., In re* Bextra & Celebrex Mktg. Sales Practices & Prod. Liab. Litig., 524 F. Supp. 2d 1166, 1172–74 (N.D. Cal. 2007); Cook v. Rockwell Int'l Corp., 580 F. Supp. 2d 1071, 1095 (D. Colo. 2006) (citing the second edition of this reference guide); *In re* W.R. Grace & Co., 355 B.R. 462, 482–83 (Bankr. D. Del. 2006).

The incidence rate of disease is defined as the number of cases of disease that develop during a specified period of time divided by the number of persons in the cohort under study.[54] Thus, the incidence rate expresses the risk that a member of the population will develop the disease within a specified period of time.

For example, a researcher studies 100 individuals who are exposed to an agent and 200 who are not exposed. After 1 year, 40 of the exposed individuals are diagnosed as having a disease, and 20 of the unexposed individuals also are diagnosed as having the disease. The relative risk of contracting the disease is calculated as follows:

- The incidence rate of disease in the exposed individuals is 40 cases per year per 100 persons (40/100), or 0.4.
- The incidence rate of disease in the unexposed individuals is 20 cases per year per 200 persons (20/200), or 0.1.
- The relative risk is calculated as the incidence rate in the exposed group (0.4) divided by the incidence rate in the unexposed group (0.1), or 4.0.

A relative risk of 4.0 indicates that the risk of disease in the exposed group is four times as high as the risk of disease in the unexposed group.[55]

In general, the relative risk can be interpreted as follows:

- If the relative risk equals 1.0, the risk in exposed individuals is the same as the risk in unexposed individuals.[56] There is no association between exposure to the agent and disease.
- If the relative risk is greater than 1.0, the risk in exposed individuals is greater than the risk in unexposed individuals. There is a positive association between exposure to the agent and the disease, which could be causal.
- If the relative risk is less than 1.0, the risk in exposed individuals is less than the risk in unexposed individuals. There is a negative association, which could reflect a protective or curative effect of the agent on risk of disease. For example, immunizations lower the risk of disease. The results suggest that immunization is associated with a decrease in disease and may have a protective effect on the risk of disease.

Although relative risk is a straightforward concept, care must be taken in interpreting it. Whenever an association is uncovered, further analysis should be

---

54. Epidemiologists also use the concept of prevalence, which measures the existence of disease in a population at a given point in time, regardless of when the disease developed. Prevalence is expressed as the proportion of the population with the disease at the chosen time. *See* Gordis, *supra* note 32, at 43–47.

55. *See* DeLuca v. Merrell Dow Pharms., Inc., 911 F.2d 941, 947 (3d Cir. 1990); Magistrini v. One Hour Martinizing Dry Cleaning, 180 F. Supp. 2d 584, 591 (D.N.J. 2002).

56. *See Magistrini*, 180 F. Supp. 2d at 591.

conducted to assess whether the association is real or a result of sampling error, confounding, or bias.[57] These same sources of error may mask a true association, resulting in a study that erroneously finds no association.

B. Odds Ratio

The odds ratio ("OR") is similar to a relative risk in that it expresses in quantitative terms the association between exposure to an agent and a disease.[58] It is a convenient way to estimate the relative risk in a case-control study when the disease under investigation is rare.[59] The odds ratio approximates the relative risk when the disease is rare.[60]

In a case-control study, the odds ratio is the ratio of the odds that a case (one with the disease) was exposed to the odds that a control (one without the disease) was exposed. In a cohort study, the odds ratio is the ratio of the odds of developing a disease when exposed to a suspected agent to the odds of developing the disease when not exposed.

Consider a case-control study, with results as shown schematically in a 2 × 2 table (Table 2):

Table 2. Cross-tabulation of cases and controls by exposure status

	Cases (with disease)	Controls (no disease)
Exposed	*a*	*b*
Not exposed	*c*	*d*

In a case-control study,

$$OR = \frac{\text{(Odds that a case was exposed)}}{\text{(Odds that a control was exposed)}}.$$

---

57. *See infra* Sections IV.B–C.

58. A relative risk cannot be calculated for a case-control study, because a case-control study begins by examining a group of persons who already have the disease. That aspect of the study design prevents a researcher from determining the rate at which individuals develop the disease. Without a rate or incidence of disease, a researcher cannot calculate a relative risk.

59. If the disease is not rare, the odds ratio is still valid to determine whether an association exists, but interpretation of its magnitude is less intuitive.

60. *See* Marcello Pagano & Kimberlee Gauvreau, Principles of Biostatistics 354 (2d ed. 2000). For further detail about the odds ratio and its calculation, see Kahn & Sempos, *supra* note 31, at 47–56.

Looking at Table 2, this ratio can be calculated as

$$\frac{(a/c)}{(b/d).}$$

This works out to *ad/bc*. Because we are multiplying two diagonal cells in the table and dividing by the product of the other two diagonal cells, the odds ratio is also called the cross-products ratio.

Consider the following hypothetical study: A researcher identifies 100 individuals with a disease who serve as "cases" and 100 people without the disease who serve as "controls" for her case-control study. Forty of the 100 cases were exposed to the agent and 60 were not. Among the control group, 20 people were exposed and 80 were not. The data can be presented in a 2 × 2 table (Table 3):

Table 3. Case-Control Study Outcome

	Cases (with disease)	Controls (no disease)
Exposed	40	20
Not exposed	60	80

The calculation of the odds ratio would be:

$$\text{OR} = \frac{(40/60)}{(20/80)} = 2.67.$$

If the disease is relatively rare in the general population (about 5% or less), the odds ratio is a good approximation of the relative risk, which means that there is almost a tripling of the disease in those exposed to the agent.[61]

---

61. The odds ratio is usually marginally greater than the relative risk. As the disease in question becomes more common, the difference between the odds ratio and the relative risk grows.

The reason why the odds ratio approximates the relative risk when the incidence of disease is small can be demonstrated by referring to Table 2. The odds ratio, as stated in the text, is *ad/bc*. The relative risk for such a study would compare the incidence of disease in the exposed group, or *a/(a + b)*, with the incidence of disease in the unexposed group or *c/(c + d)*. The relative risk would be:

$$\frac{a/\left(a+b\right)}{c/\left(c+d\right)} = \frac{a/\left(c+d\right)}{c/\left(a+b\right)}$$

When the incidence of disease is low, *a* and *c* will be small in relation to *b* and *d*, and the relative risk will then approximate the odds ratio of *ad/bc*. *See* Leon Gordis, Epidemiology 208–09 (4th ed. 2009).

C. Attributable Risk

A frequently used measurement of risk is the attributable risk ("AR"). The attributable risk represents the amount of disease among exposed individuals that can be attributed to the exposure. It also can be expressed as the proportion of the disease among exposed individuals that is associated with the exposure (also called the "attributable proportion of risk," the "etiologic fraction," or the "attributable risk percent"). The attributable risk reflects the maximum proportion of the disease that can be attributed to exposure to an agent and consequently the maximum proportion of disease that could be potentially prevented by blocking the effect of the exposure or by eliminating the exposure.[62] In other words, if the association is causal, the attributable risk is the proportion of disease in an exposed population that might be caused by the agent and that might be prevented by eliminating exposure to that agent (see Figure 3).[63]

Figure 3. Risks in exposed and unexposed groups.
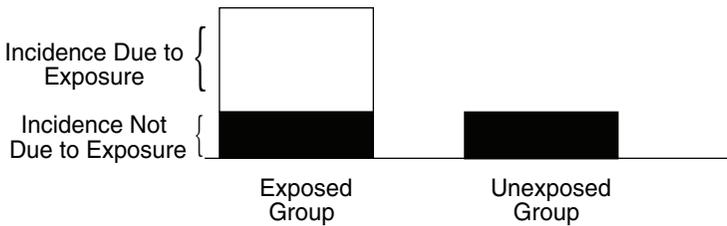
To determine the proportion of a disease that is attributable to an exposure, a researcher would need to know the incidence of the disease in the exposed group and the incidence of disease in the unexposed group. The attributable risk is

$$AR = \frac{(\text{incidence in the exposed}) - (\text{incidence in the unexposed})}{\text{incidence in the exposed}}$$

---

62. Kenneth J. Rothman et al., Modern Epidemiology 297 (3d ed. 2008); *see also* Landrigan v. Celotex Corp., 605 A.2d 1079, 1086 (N.J. 1992) (illustrating that a relative risk of 1.55 conforms to an attributable risk of 35%, that is, $(1.55 - 1.0)/1.55 = .35$, or 35%).

63. Risk is not zero for the control group (those not exposed) when there are other causal chains that cause the disease that do not require exposure to the agent. For example, some birth defects are the result of genetic sources, which do not require the presence of any environmental agent. Also, some degree of risk in the control group may be the result of background exposure to the agent being studied. For example, nonsmokers in a control group may have been exposed to passive cigarette smoke, which is responsible for some cases of lung cancer and other diseases. *See also* Ethyl Corp. v. EPA, 541 F.2d 1, 25 (D.C. Cir. 1976). There are some diseases that do not occur without exposure to an agent; these are known as signature diseases. *See infra* note 177.

The attributable risk can be calculated using the example described in Section III.A. Suppose a researcher studies 100 individuals who are exposed to a substance and 200 who are not exposed. After 1 year, 40 of the exposed individuals are diagnosed as having a disease, and 20 of the unexposed individuals are also diagnosed as having the disease.

- The incidence of disease in the exposed group is 40 persons out of 100 who contract the disease in a year.
- The incidence of disease in the unexposed group is 20 persons out of 200 (or 10 out of 100) who contract the disease in a year.
- The proportion of disease that is attributable to the exposure is 30 persons out of 40, or 75%.

This means that 75% of the disease in the exposed group is attributable to the exposure. We should emphasize here that "attributable" does not necessarily mean "caused by." Up to this point, we have only addressed associations. Inferring causation from an association is addressed in Section V.

D. *Adjustment for Study Groups That Are Not Comparable*

Populations often differ in characteristics that relate to disease risk, such as age, sex, and race. Those who live in Florida have a much higher death rate than those who live in Alaska.[64] Is sunshine dangerous? Perhaps, but the Florida population is much older than the Alaska population, and some adjustment must be made for the differences in age distribution in the two states in order to compare disease or death rates between populations. The technique used to accomplish this is called adjustment, and two types of adjustment are used—direct and indirect. In direct adjustment (e.g., when based on age), overall disease/death rates are calculated for each population as though each had the age distribution of another standard, or reference, population, using the age-specific disease/death rates for each study population. We can then compare these overall rates, called age-adjusted rates, knowing that any difference between these rates cannot be attributed to differences in age, since both age-adjusted rates were generated using the same standard population.

Indirect adjustment is used when the age-specific rates for a study population are not known. In that case, the overall disease/death rate for the standard/reference population is recalculated based on the age distribution of the population of interest using the age-specific rates of the standard population. Then, the actual number of disease cases/deaths in the population of interest can be compared with

---

64. *See* Lilienfeld & Stolley, *supra* note 35, at 68–70 (the mortality rate in Florida is approximately three times what it is in Alaska).

the number in the reference population that would be expected if the reference population had the age distribution of the population of interest.

This ratio is called the standardized mortality ratio (SMR). When the outcome of interest is disease rather than death, it is called the standardized morbidity ratio.[65] If the ratio equals 1.0, the observed number of deaths equals the expected number of deaths, and the mortality rate of the population of interest is no different from that of the reference population. If the SMR is greater than 1.0, the population of interest has a higher mortality risk than that of the reference population, and if the SMR is less than 1.0, the population of interest has a lower mortality rate than that of the reference population.

Thus, age adjustment provides a way to compare populations while in effect holding age constant. Adjustment is used not only for comparing mortality rates in different populations but also for comparing rates in different groups of subjects selected for study in epidemiologic investigations. Although this discussion has focused on adjusting for age, it is also possible to adjust for any number of other variables, such as gender, race, occupation, and socioeconomic status. It is also possible to adjust for several factors simultaneously.[66]

IV. What Sources of Error Might Have Produced a False Result?

Incorrect study results occur in a variety of ways. A study may find a positive association (relative risk greater than 1.0) when there is no true association. Or a study may erroneously result in finding that that there is no association when in reality there is. A study may also find an association when one truly exists, but the association found may be greater or less than the real association.

Three general categories of phenomena can result in an association found in a study to be erroneous: chance, bias, and confounding. Before any inferences about causation are drawn from a study, the possibility of these phenomena must be examined.[67]

---

65. *See* Taylor v. Airco, Inc., 494 F. Supp. 2d 21, 25 n.4 (D. Mass. 2007) (explaining SMR and its relationship with relative risk). For an example of adjustment used to calculate an SMR for workers exposed to benzene, see Robert A. Rinsky et al., *Benzene and Leukemia: An Epidemiologic Risk Assessment,* 316 New Eng. J. Med. 1044 (1987).

66. For further elaboration on adjustment, see Gordis, *supra* note 32, at 73–78; Philip Cole, *Causality in Epidemiology, Health Policy, and Law,* 27 Envtl. L. Rep. 10,279, 10,281 (1997).

67. *See* Cole, *supra* note 65, at 10,285. In *DeLuca v. Merrell Dow Pharmaceuticals, Inc.,* 911 F.2d 941, 955 (3d Cir. 1990), the court recognized and discussed random sampling error. It then went on to refer to other errors (e.g., systematic bias) that create as much or more error in the outcome of a study. For a similar description of error in study procedure and random sampling, see David H. Kaye & David A. Freedman, Reference Guide on Statistics, Section IV, in this manual.

The findings of a study may be the result of chance (or random error). In designing a study, the size of the sample can be increased to reduce (but not eliminate) the likelihood of random error. Once a study has been completed, statistical methods (discussed in Section IV.A) permit an assessment of the extent to which the results of a study may be due to random error.

The two main techniques for assessing random error are statistical significance and confidence intervals. A study that is statistically significant has results that are unlikely to be the result of random error, although any criterion for "significance" is somewhat arbitrary. A confidence interval provides both the relative risk (or other risk measure) found in the study and a range (interval) within which the risk likely would fall if the study were repeated numerous times. These two techniques (which are closely related) are explained in Section IV.A.

We should emphasize a matter that those unfamiliar with statistical methodology frequently find confusing: That a study's results are statistically significant says nothing about the importance of the magnitude of any association (i.e., the relative risk or odds ratio) found in a study or about the biological or clinical importance of the finding.[68] "Significant," as used with the adjective "statistically," does not mean important. A study may find a statistically significant relationship that is quite modest—perhaps it increases the risk only by 5%, which is equivalent to a relative risk of 1.05.[69] An association may be quite large—the exposed cohort might be 10 times more likely to develop disease than the control group—but the association is not statistically significant because of the potential for random error given a small sample size. In short, *statistical significance is not about the size of the risk found in a study*.

Bias (or systematic error) also can produce error in the outcome of a study. Epidemiologists attempt to minimize bias through their study design, including data collection protocols. Study designs are developed before they begin gathering data. However, even the best designed and conducted studies have biases, which may be subtle. Consequently, after data collection is completed, analytical tools are often used to evaluate potential sources of bias. Sometimes, after bias is identified, the epidemiologist can determine whether the bias would tend to inflate or dilute any association that may exist. Identification of the bias may permit the

---

68. *See* Modern Scientific Evidence, *supra* note 2, § 6.36 at 358 ("Statisticians distinguish between 'statistical' and 'practical' significance. . . ."); Cole, *supra* note 65, at 10,282. Understandably, some courts have been confused about the relationship between statistical significance and the magnitude of the association. *See* Hyman & Armstrong, P.S.C. v. Gunderson, 279 S.W.3d 93, 102 (Ky. 2008) (describing a small increased risk as being considered statistically insignificant and a somewhat larger risk as being considered statistically significant.); *In re* Pfizer Inc. Sec. Litig., 584 F. Supp. 2d 621, 634–35 (S.D.N.Y. 2008) (confusing the magnitude of the effect with whether the effect was statistically significant); *In re* Joint E. & S. Dist. Asbestos Litig., 827 F. Supp. 1014, 1041 (S.D.N.Y. 1993) (concluding that any relative risk less than 1.50 is statistically insignificant), *rev'd on other grounds,* 52 F.3d 1124 (2d Cir. 1995).

69. In general, small effects that are statistically significant require larger sample sizes. When effects are larger, generally fewer subjects are required to produce statistically significant findings.

epidemiologist to make an assessment of whether the study's conclusions are valid. Epidemiologists may reanalyze a study's data to correct for a bias identified in a completed study or to validate the analytical methods used.[70] Common biases and how they may produce invalid results are described in Section IV.B.

Finally, a study may reach incorrect conclusions about causation because, although the agent and disease are associated, the agent is not a true causal factor. Rather, the agent may be associated with another agent that is the true causal factor, and this latter factor confounds the relationship being examined in the study. Confounding is explained in Section IV.C.

A. What Statistical Methods Exist to Evaluate the Possibility of Sampling Error?[71]

Before detailing the statistical methods used to assess random error (which we use as synonymous with sampling error), two concepts are explained that are central to epidemiology and statistical analysis. Understanding these concepts should facilitate comprehension of the statistical methods.

Epidemiologists often refer to the true association (also called "real association"), which is the association that really exists between an agent and a disease and that might be found by a perfect (but nonexistent) study. The true association is a concept that is used in evaluating the results of a given study even though its value is unknown. By contrast, a study's outcome will produce an observed association, which is known.

Formal procedures for statistical testing begin with the null hypothesis, which posits that there is no true association (i.e., a relative risk of 1.0) between the agent and disease under study. Data are gathered and analyzed to see whether they disprove[72] the null hypothesis. The data are subjected to statistical testing to assess the plausibility that any association found is a result of random error or whether it supports rejection of the null hypothesis. The use of the null hypothesis for this testing should not be understood as the a priori belief of the investigator. When epidemiologists investigate an agent, it is usually because they hypothesize that the agent is a cause of some outcome. Nevertheless, epidemiologists prepare their

---

70. *E.g.,* Richard A. Kronmal et al., *The Intrauterine Device and Pelvic Inflammatory Disease: The Women's Health Study Reanalyzed,* 44 J. Clin. Epidemiol. 109 (1991) (a reanalysis of a study that found an association between the use of IUDs and pelvic inflammatory disease concluded that IUDs do not increase the risk of pelvic inflammatory disease).

71. For a bibliography on the role of statistical significance in legal proceedings, see Sanders, *supra* note 13, at 329 n.138.

72. *See, e.g.,* Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579, 593 (1993) (scientific methodology involves generating and testing hypotheses).

study designs and test the plausibility that any association found in a study was the result of random error by using the null hypothesis.[73]

1. *False positives and statistical significance*

When a study results in a positive association (i.e., a relative risk greater than 1.0), epidemiologists try to determine whether that outcome represents a true association or is the result of random error.[74] Random error is illustrated by a fair coin (i.e., not modified to produce more heads than tails [or vice versa]). On average, for example, we would expect that coin tosses would yield half heads and half tails. But sometimes, a set of coin tosses might yield an unusual result, for example, six heads out of six tosses,[75] an occurrence that would result, purely by chance, in less than 2% of a series of six tosses. In the world of epidemiology, sometimes the study findings, merely by chance, do not reflect the true relationships between an agent and outcome. Any single study—even a clinical trial—is in some ways analogous to a set of coin tosses, being subject to the play of chance. Thus, for example, even though the true relative risk (in the total population) is 1.0, an epidemiologic study of a particular study population may find a relative risk greater than (or less

73. *See* DeLuca v. Merrell Dow Pharms., Inc., 911 F.2d 941, 945 (3d Cir. 1990); United States v. Philip Morris USA, Inc., 449 F. Supp. 2d 1, 706 n.29 (D.D.C. 2006); Stephen E. Fienberg et al., *Understanding and Evaluating Statistical Evidence in Litigation,* 36 Jurimetrics J. 1, 21–24 (1995).

74. Hypothesis testing is one of the most counterintuitive techniques in statistics. Given a set of epidemiologic data, one wants to ask the straightforward, obvious question: What is the probability that the difference between two samples reflects a real difference between the populations from which they were taken? Unfortunately, there is no way to answer this question directly or to calculate the probability. Instead, statisticians—and epidemiologists—address a related but very different question: If there really is no difference between the populations, how probable is it that one would find a difference at least as large as the observed difference between the samples? *See* Modern Scientific Evidence, *supra* note 2, § 6:36, at 359 ("it is easy to mistake the *p*-value for the probability that there is no difference"); Expert Evidence: A Practitioner's Guide to Law, Science, and the FJC Manual 91 (Bert Black & Patrick W. Lee eds., 1997). Thus, the *p*-value for a given study does not provide a rate of error or even a probability of error for an epidemiologic study. In *Daubert v. Merrell Dow Pharmaceuticals, Inc.,* 509 U.S. 579, 593 (1993), the Court stated that "the known or potential rate of error" should ordinarily be considered in assessing scientific reliability. Epidemiology, however, unlike some other methodologies—fingerprint identification, for example—does not permit an assessment of its accuracy by testing with a known reference standard. A *p*-value provides information only about the plausibility of random error given the study result, but the true relationship between agent and outcome remains unknown. Moreover, a *p*-value provides no information about whether other sources of error—bias and confounding—exist and, if so, their magnitude. In short, for epidemiology, there is no way to determine a rate of error. *See* Kumho Tire Co. v. Carmichael, 526 U.S. 137, 151 (1999) (recognizing that for different scientific and technical inquiries, different considerations will be appropriate for assessing reliability); Cook v. Rockwell Int'l Corp., 580 F. Supp. 2d 1071, 1100 (D. Colo. 2006) ("Defendants have not argued or presented evidence that . . . a method by which an overall 'rate of error' can be calculated for an epidemiologic study.")

75. *DeLuca*, 911 F.2d at 946–47.

than) 1.0 because of random error or chance.[76] An erroneous conclusion that the null hypothesis is false (i.e., a conclusion that there is a difference in risk when no difference actually exists) owing to random error is called a false–positive error (also Type I error or alpha error).

Common sense leads one to believe that a large enough sample of individuals must be studied if the study is to identify a relationship between exposure to an agent and disease that truly exists. Common sense also suggests that by enlarging the sample size (the size of the study group), researchers can form a more accurate conclusion and reduce the chance of random error in their results. Both statements are correct and can be illustrated by a test to determine if a coin is fair. A test in which a fair coin is tossed 1000 times is more likely to produce close to 50% heads than a test in which the coin is tossed only 10 times. It is far more likely that a test of a fair coin with 10 tosses will come up, for example, with 80% heads than will a test with 1000 tosses. With large numbers, the outcome of the test is less likely to be influenced by random error, and the researcher would have greater confidence in the inferences drawn from the data.[77]

One means for evaluating the possibility that an observed association could have occurred as a result of random error is by calculating a *p*-valu*e*.[78] A *p*-value represents the probability that an observed positive association could result from random error even if no association were in fact present. Thus, a *p*-value of .1 means that there is a 10% chance that values at least as large as the observed relative risk could have occurred by random error, with no association actually present in the population.[79]

To minimize false positives, epidemiologists use a convention that the *p*-value must fall below some selected level known as alpha or significance level for the results of the study to be statistically significant.[80] Thus, an outcome is statistically significant when the observed *p*-value for the study falls below the preselected

---

76. *See* Magistrini v. One Hour Martinizing Dry Cleaning, 180 F. Supp. 2d 584, 592 (D.N.J. 2002) (citing the second edition of this reference guide).

77. This explanation of numerical stability was drawn from Brief for Professor Alvan R. Feinstein as Amicus Curiae Supporting Respondents at 12–13, Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579 (1993) (No. 92-102). *See also* Allen v. United States, 588 F. Supp. 247, 417–18 (D. Utah 1984), *rev'd on other grounds,* 816 F.2d 1417 (10th Cir. 1987). The *Allen* court observed that although "[s]mall communities or groups of people are deemed 'statistically unstable'" and "data from small populations must be handled with care [, it] does not mean that [the data] cannot provide substantial evidence in aid of our effort to describe and understand events."

78. *See also* David H. Kaye & David A. Freedman, Reference Guide on Statistics, Section IV.B, in this manual (the *p*-value reflects the implausibility of the null hypothesis).

79. Technically, a *p*-value of .1 means that if in fact there is no association, 10% of all similar studies would be expected to yield an association the same as, or greater than, the one found in the study due to random error.

80. Cook v. Rockwell Int'l Corp., 580 F. Supp. 2d 1071, 1100–01 (D. Colo. 2006) (discussing *p*-values and their relationship with statistical significance); *Allen*, 588 F. Supp. at 416–17 (discussing statistical significance and selection of a level of alpha); *see also* Sanders, *supra* note 13, at 343–44 (explaining alpha, beta, and their relationship to sample size); *Developments in the Law—Confronting*

significance level. The most common significance level, or alpha, used in science is .05.[81] A .05 value means that the probability is 5% of observing an association at least as large as that found in the study when in truth there is no association.[82] Although .05 is often the significance level selected, other levels can and have been used.[83] Thus, in its study of the effects of second-hand smoke, the U.S.

---

the New Challenges of Scientific Evidence, 108 Harv. L. Rev. 1481, 1535–36, 1540–46 (1995) [hereafter *Developments in the Law*].

81. A common error made by lawyers, judges, and academics is to equate the level of alpha with the legal burden of proof. Thus, one will often see a statement that using an alpha of .05 for statistical significance imposes a burden of proof on the plaintiff far higher than the civil burden of a preponderance of the evidence (i.e., greater than 50%). *See, e.g., In re* Ephedra Prods. Liab. Litig., 393 F. Supp. 2d 181, 193 (S.D.N.Y. 2005); Marmo v. IBP, Inc., 360 F. Supp. 2d 1019, 1021 n.2 (D. Neb. 2005) (an expert toxicologist who stated that science requires proof with 95% certainty while expressing his understanding that the legal standard merely required more probable than not). *But see* Giles v. Wyeth, Inc., 500 F. Supp. 2d 1048, 1056–57 (S.D. Ill. 2007) (quoting the second edition of this reference guide).

Comparing a selected *p*-value with the legal burden of proof is mistaken, although the reasons are a bit complex and a full explanation would require more space and detail than is feasible here. Nevertheless, we sketch out a brief explanation: First, alpha does not address the likelihood that a plaintiff's disease was caused by exposure to the agent; the magnitude of the association bears on that question. *See infra* Section VII. Second, significance testing only bears on whether the observed magnitude of association arose as a result of random chance, not on whether the null hypothesis is true. Third, using stringent significance testing to avoid false-positive error comes at a complementary cost of inducing false-negative error. Fourth, using an alpha of .5 would not be equivalent to saying that the probability the association found is real is 50%, and the probability that it is a result of random error is 50%. Statistical methodology does not permit assessments of those probabilities. *See* Green, *supra* note 47, at 686; Michael D. Green, *Science Is to Law as the Burden of Proof Is to Significance Testing,* 37 Jurimetrics J. 205 (1997) (book review); *see also* David H. Kaye, *Apples and Oranges: Confidence Coefficients and the Burden of Persuasion,* 73 Cornell L. Rev. 54, 66 (1987); David H. Kaye & David A. Freedman, Reference Guide on Statistics, Section IV.B.2, in this manual; Turpin v. Merrell Dow Pharms., Inc., 959 F.2d 1349, 1357 n.2 (6th Cir. 1992), *cert. denied,* 506 U.S. 826 (1992); *cf. DeLuca,* 911 F.2d at 959 n.24 ("The relationship between confidence levels and the more likely than not standard of proof is a very complex one . . . and in the absence of more education than can be found in this record, we decline to comment further on it.").

82. This means that if one conducted an examination of a large number of associations in which the true RR equals 1, on average 1 in 20 associations found to be statistically significant at a .05 level would be spurious. When researchers examine many possible associations that might exist in their data—known as data dredging—we should expect that even if there are no true causal relationships, those researchers will find statistically significant associations in 1 of every 20 associations examined. *See* Rachel Nowak, *Problems in Clinical Trials Go Far Beyond Misconduct,* 264 Sci. 1538, 1539 (1994).

83. A significance test can be either one-tailed or two-tailed, depending on the null hypothesis selected by the researcher. Because most investigators of toxic substances are only interested in whether the agent increases the incidence of disease (as distinguished from providing protection from the disease), a one-tailed test is often viewed as appropriate. *In re* Phenylpropanolamine (PPA) Prods. Liab. Litig., 289 F. Supp. 2d 1230, 1241 (W.D. Wash. 2003) (accepting the propriety of a one-tailed test for statistical significance in a toxic substance case); United States v. Philip Morris USA, Inc., 449 F. Supp. 2d 1, 701 (D.D.C. 2006) (explaining the basis for EPA's decision to use one-tailed test in assessing whether second-hand smoke was a carcinogen). *But see* Good v. Fluor Daniel Corp., 222 F. Supp. 2d 1236, 1243 (E.D. Wash. 2002). For an explanation of the difference

Environmental Protection Agency (EPA) used a .10 standard for significance testing.[84]

There is some controversy among epidemiologists and biostatisticians about the appropriate role of significance testing.[85] To the strictest significance testers,

between one–tailed and two–tailed tests, see David H. Kaye & David A. Freedman, Reference Guide on Statistics, Section IV.C.2, in this manual.

84. U.S. Environmental Protection Agency, Respiratory Health Effects of Passive Smoking: Lung Cancer and Other Disorders (1992); *see also Turpin,* 959 F.2d at 1353–54 n.1 (confidence level frequently set at 95%, although 90% (which corresponds to an alpha of .10) is also used; selection of the value is "somewhat arbitrary").

85. Similar controversy exists among the courts that have confronted the issue of whether statistically significant studies are required to satisfy the burden of production. The leading case advocating statistically significant studies is *Brock v. Merrell Dow Pharmaceuticals, Inc.,* 874 F.2d 307, 312 (5th Cir. 1989), *amended,* 884 F.2d 167 (5th Cir.), *cert. denied,* 494 U.S. 1046 (1990). Overturning a jury verdict for the plaintiff in a Bendectin case, the court observed that no statistically significant study had been published that found an increased relative risk for birth defects in children whose mothers had taken Bendectin. The court concluded: "[W]e do not wish this case to stand as a bar to future Bendectin cases in the event that new and statistically significant studies emerge which would give a jury a firmer basis on which to determine the issue of causation." *Brock,* 884 F.2d at 167.

A number of courts have followed the *Brock* decision or have indicated strong support for significance testing as a screening device. *See* Good v. Fluor Daniel Corp., 222 F. Supp. 2d 1236, 1243 (E.D. Wash. 2002) ("In the absence of a statistically significant difference upon which to opine, Dr. Au's opinion must be excluded under *Daubert.*"); Miller v. Pfizer, Inc., 196 F. Supp. 2d 1062, 1080 (D. Kan. 2002) (the expert must have statistically significant studies to serve as basis of opinion on causation); Kelley v. Am. Heyer-Schulte Corp., 957 F. Supp. 873, 878 (W.D. Tex. 1997) (the lower end of the confidence interval must be above 1.0—equivalent to requiring that a study be statistically significant—before a study may be relied upon by an expert), *appeal dismissed,* 139 F.3d 899 (5th Cir. 1998); Renaud v. Martin Marietta Corp., 749 F. Supp. 1545, 1555 (D. Colo. 1990) (quoting *Brock* approvingly), *aff'd,* 972 F.2d 304 (10th Cir. 1992).

By contrast, a number of courts are more cautious about or reject using significance testing as a necessary condition, instead recognizing that assessing the likelihood of random error is important in determining the probative value of a study. In *Allen v. United States,* 588 F. Supp. 247, 417 (D. Utah 1984), the court stated, "The cold statement that a given relationship is not 'statistically significant' cannot be read to mean there is no probability of a relationship." The Third Circuit described confidence intervals (i.e., the range of values that would be found in similar studies due to chance, with a specified level of confidence) and their use as an alternative to statistical significance in *DeLuca v. Merrell Dow Pharmaceuticals, Inc.,* 911 F.2d 941, 948–49 (3d Cir. 1990). *See also* Milward v. Acuity Specialty Products Group, Inc., 639 F.3d 11, 24-25 (1st Cir. 2011) (recognizing the difficulty of obtaining statistically significant results when the disease under investigation occurs rarely and concluding that district court erred in imposing a statistical significance threshold); Turpin v. Merrell Dow Pharms., Inc., 959 F.2d 1349, 1357 (6th Cir. 1992) ("The defendant's claim overstates the persuasive power of these statistical studies. An analysis of this evidence demonstrates that it is possible that Bendectin causes birth defects even though these studies do not detect a significant association."); *In re* Viagra Prods. Liab. Litig., 572 F. Supp. 2d 1071, 1090 (D. Minn. 2008) (holding that, for purposes of supporting an opinion on general causation, a study does not have to find results with statistical significance); United States v. Philip Morris USA, Inc., 449 F. Supp. 2d 1, 706 n.29 (D.D.C. 2006) (rejecting the position of an expert who denied that the causal connection between smoking and lung cancer had been established, in part, on the ground that any study that found an association that was not statistically significant must be excluded from consideration); Cook v. Rockwell Int'l Corp., 580 F. Supp.

any study whose *p*-value is not less than the level chosen for statistical significance should be rejected as inadequate to disprove the null hypothesis. Others are critical of using strict significance testing, which rejects all studies with an observed *p*-value below that specified level. Epidemiologists have become increasingly sophisticated in addressing the issue of random error and examining the data from a study to ascertain what information they may provide about the relationship between an agent and a disease, without the necessity of rejecting all studies that are not statistically significant.[86] Meta-analysis, as well, a method for pooling the results of multiple studies, sometimes can ameliorate concerns about random error.[87]

Calculation of a confidence interval permits a more refined assessment of appropriate inferences about the association found in an epidemiologic study.[88]

2d 1071, 1103 (D. Colo. 2006) ("The statistical significance or insignificance of Dr. Clapp's results may affect the weight given to his testimony, but does not determine its admissibility under Rule 702."); *In re* Ephedra Prods. Liab. Litig., 393 F. Supp. 2d 181, 186 (S.D.N.Y. 2005) ("[T]he absence of epidemiologic studies establishing an increased risk from ephedra of sufficient statistical significance to meet scientific standards of causality does not mean that the causality opinions of the PCC's experts must be excluded entirely.").

Although the trial court had relied in part on the absence of statistically significant epidemiologic studies, the Supreme Court in *Daubert v. Merrell Dow Pharmaceuticals, Inc.,* 509 U.S. 579 (1993), did not explicitly address the matter. The Court did, however, refer to "the known or potential rate of error" in identifying factors relevant to the scientific validity of an expert's methodology. *Id.* at 594. The Court did not address any specific rate of error, although two cases that it cited affirmed the admissibility of voice spectrograph results that the courts reported were subject to a 2%–6% chance of error owing to either false matches or false eliminations. One commentator has concluded, "*Daubert* did not set a threshold level of statistical significance either for admissibility or for sufficiency of scientific evidence." *Developments in the Law, supra* note 79, at 1535–36, 1540–46. The Supreme Court in *General Electric Co. v. Joiner,* 522 U.S. 136, 145–47 (1997), adverted to the lack of statistical significance in one study relied on by an expert as a ground for ruling that the district court had not abused its discretion in excluding the expert's testimony.

In *Matrixx Initiatives, Inc. v. Siracusano,* 131 S. Ct. 1309 (2011), the Supreme Court was confronted with a question somewhat different from the relationship between statistically significant study results and causation. *Matrixx* was a securities fraud case in which the defendant argued that unless adverse event reports from use of a drug are statistically significant, the information about them is not material, as a matter of law (materiality is required as an element of a fraud claim). Defendant's claim was premised on the idea that only statistically significant results can be a basis for an inference of causation. The Court, unanimously, rejected that claim, citing cases in which courts had permitted expert witnesses to testify to toxic causation in the absence of any statistically significant studies.

For a hypercritical assessment of statistical significance testing that nevertheless identifies much inappropriate overreliance on it, see Stephen T. Ziliak & Deidre N. McCloskey, The Cult of Statistical Significance (2008).

86. *See* Sanders, *supra* note 13, at 342 (describing the improved handling and reporting of statistical analysis in studies of Bendectin after 1980).

87. *See infra* Section VI.

88. Kenneth Rothman, Professor of Public Health at Boston University and Adjunct Professor of Epidemiology at the Harvard School of Public Health, is one of the leaders in advocating the use of confidence intervals and rejecting strict significance testing. In *DeLuca,* 911 F.2d at 947, the Third Circuit discussed Rothman's views on the appropriate level of alpha and the use of con-

A confidence interval is a range of possible values calculated from the results of a study. If a 95% confidence interval is specified, the range encompasses the results we would expect 95% of the time if samples for new studies were repeatedly drawn from the same population. Thus, the width of the interval reflects random error.

The narrower the confidence interval, the more statistically stable the results of the study. The advantage of a confidence interval is that it displays more information than significance testing. "Statistically significant" does not convey the magnitude of the association found in the study or indicate how statistically stable that association is. A confidence interval shows the boundaries of the relative risk based on selected levels of alpha or statistical significance. Just as the *p*-value does not provide the probability that the risk estimate found in a study is correct, the confidence interval does not provide the range within which the true risk must lie. Rather, the confidence interval reveals the likely range of risk estimates consistent with random error. An example of two confidence intervals that might be calculated for a given relative risk is displayed in Figure 4.

Figure 4. Confidence intervals.
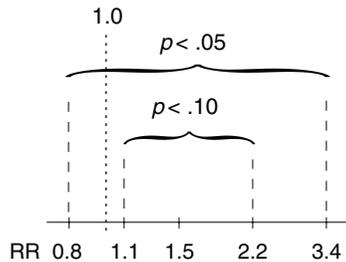
The confidence intervals shown in Figure 4 are for a study that found a relative risk of 1.5, with boundaries of 0.8 to 3.4 when the alpha is set to .05 (equivalently, a confidence level of .95), and with boundaries of 1.1 to 2.2 when alpha is set to .10 (equivalently, a confidence level of .90). The confidence interval for alpha equal to .10 is narrower because it encompasses only 90% of the expected test results. By contrast, the confidence interval for alpha equal to .05 includes the expected outcomes for 95% of the tests. To generalize this point, the lower the alpha chosen (and therefore the more stringent the exclusion of possible random error) the wider the confidence interval. At a given alpha, the width of the confidence interval is

fidence intervals. In *Turpin*, 959 F.2d at 1353–54 n.1, the court discussed the relationship among confidence intervals, alpha, and power. *See also* Cook v. Rockwell Int'l Corp., 580 F. Supp. 2d 1071, 1100–01 (D. Colo. 2006) (discussing confidence intervals, alpha, and significance testing). The use of confidence intervals in evaluating sampling error more generally than in the epidemiologic context is discussed in David H. Kaye & David A. Freedman, Reference Guide on Statistics, Section IV.A, in this manual.

determined by sample size. All other things being equal, the larger the sample size, the narrower the confidence boundaries (indicating greater numerical stability). For a given risk estimate, a narrower confidence interval reflects a decreased likelihood that the association found in the study would occur by chance if the true association is 1.0.[89]

For the example in Figure 4, the boundaries of the confidence interval with alpha set at .05 encompass a relative risk of 1.0, and the result would be said to be not statistically significant at the .05 level. Alternatively, if the confidence boundaries are defined as an alpha equal to .10, then the confidence interval no longer includes a relative risk of 1.0, and the result would be characterized as statistically significant at the .10 level.

2. False negatives

As Figure 4 illustrates, false positives can be reduced by adopting more stringent values for alpha. Using an alpha of .05 will result in fewer false positives than using an alpha of .10, and an alpha of .01 or .001 would produce even fewer false positives.[90] The tradeoff for reducing false positives is an increase in false-negative errors (also called beta errors or Type II errors). This concept reflects the possibility that a study will be interpreted as "negative" (not disproving the null

---

89. Where multiple epidemiologic studies are available, a technique known as meta-analysis (*see infra* Section VI) may be used to combine the results of the studies to reduce the numerical instability of all the studies. *See generally* Diana B. Petitti, Meta-analysis, Decision Analysis, and Cost-Effectiveness Analysis: Methods for Quantitative Synthesis in Medicine (2d ed. 2000). Meta-analysis is better suited to combining results from randomly controlled experimental studies, but if carefully performed it may also be helpful for observational studies, such as those in the epidemiologic field. *See* Zachary B. Gerberg & Ralph I. Horwitz, *Resolving Conflicting Clinical Trials: Guidelines for Meta-Analysis,* 41 J. Clin. Epidemiol. 503 (1988). In *In re Bextra & Celebrex Marketing Sales Practices & Products Liability Litigation,* 524 F. Supp. 2d 1166 (N.D. Cal. 2007), the court relied on several meta-analyses of Celebrex at a 200-mg dose to conclude that the plaintiffs' experts who proposed to testify to toxicity at that dosage failed to meet the requirements of *Daubert.* The court criticized those experts for the wholesale rejection of meta-analyses of observational studies.

In *In re Paoli Railroad Yard PCB Litigation,* 916 F.2d 829, 856–57 (3d Cir. 1990), the court discussed the use and admissibility of meta-analysis as a scientific technique. Overturning the district court's exclusion of a report using meta-analysis, the Third Circuit observed that meta-analysis is a regularly used scientific technique. The court recognized that the technique might be poorly performed, and it required the district court to reconsider the validity of the expert's work in performing the meta-analysis. *See also* E.R. Squibb & Sons, Inc. v. Stuart Pharms., No. 90-1178, 1990 U.S. Dist. LEXIS 15788, at ★41 (D.N.J. Oct. 16, 1990) (acknowledging the utility of meta-analysis but rejecting its use in that case because one of the two studies included was poorly performed); Tobin v. Astra Pharm. Prods., Inc., 993 F.2d 528, 538–39 (6th Cir. 1992) (identifying an error in the performance of a meta-analysis, in which the Food and Drug Administration pooled data from control groups in different studies in which some gave the controls a placebo and others gave the controls an alternative treatment).

90. It is not uncommon in genome-wide association studies to set the alpha at .00001 or even lower because of the large number of associations tested in such studies. Reducing alpha is designed to limit the number of false-positive findings.

hypothesis), when in fact there is a true association of a specified magnitude.[91] The beta for any study can be calculated only based on a specific alternative hypothesis about a given positive relative risk and a specific level of alpha selected.[92]

3. Power

When a study fails to find a statistically significant association, an important question is whether the result tends to exonerate the agent's toxicity or is essentially inconclusive with regard to toxicity.[93] The concept of power can be helpful in evaluating whether a study's outcome is exonerative or inconclusive.[94]

The power of a study is the probability of finding a statistically significant association of a given magnitude (if it exists) in light of the sample sizes used in the study. The power of a study depends on several factors: the sample size; the level of alpha (or statistical significance) specified; the background incidence of disease; and the specified relative risk that the researcher would like to detect.[95] Power curves can be constructed that show the likelihood of finding any given relative risk in light of these factors. Often, power curves are used in the design of a study to determine what size the study populations should be.[96]

The power of a study is the complement of beta $(1 - \beta)$. Thus, a study with a likelihood of .25 of failing to detect a true relative risk of 2.0[97] or greater has a power of .75. This means the study has a 75% chance of detecting a true relative risk of 2.0. If the power of a negative study to find a relative risk of 2.0 or greater

---

91. *See also* DeLuca v. Merrell Dow Pharms., Inc., 911 F.2d 941, 947 (3d Cir. 1990).

92. *See* Green, *supra* note 47, at 684–89.

93. Even when a study or body of studies tends to exonerate an agent, that does not establish that the agent is absolutely safe. *See* Cooley v. Lincoln Elec. Co., 693 F. Supp. 2d 767 (N.D. Ohio 2010). Epidemiology is not able to provide such evidence.

94. *See* Fienberg et al., *supra* note 72, at 22–23. Thus, in *Smith v. Wyeth-Ayerst Labs. Co.,* 278 F. Supp. 2d 684, 693 (W.D.N.C. 2003) and *Cooley v. Lincoln Electric Co.,* 693 F. Supp. 2d 767, 773 (N.D. Ohio 2010), the courts recognized that the power of a study was critical to assessing whether the failure of the study to find a statistically significant association was exonerative of the agent or inconclusive. *See also* Procter & Gamble Pharms., Inc. v. Hoffmann-LaRoche Inc., No. 06 Civ. 0034(PAC), 2006 WL 2588002, at *32 n.16 (S.D.N.Y. Sept. 6, 2006) (discussing power curves and quoting the second edition of this reference guide); *In re* Phenylpropanolamine (PPA) Prods. Liab. Litig., 289 F. Supp. 2d 1230, 1243–44 (W.D. Wash. 2003) (explaining expert's testimony that "statistical reassurance as to lack of an effect would require an upper bound of a reasonable confidence interval close to the null value"); Ruff v. Ensign-Bickford Indus., Inc., 168 F. Supp. 2d 1271, 1281 (D. Utah 2001) (explaining why a study should be treated as inconclusive rather than exonerative based on small number of subjects in study).

95. *See* Malcolm Gladwell, *How Safe Are Your Breasts?* New Republic, Oct. 24, 1994, at 22, 26.

96. For examples of power curves, see Kenneth J. Rothman, Modern Epidemiology 80 (1986); Pagano & Gauvreau, *supra* note 59, at 245.

97. We use a relative risk of 2.0 for illustrative purposes because of the legal significance courts have attributed to this magnitude of association. *See infra* Section VII.

is low, it has substantially less probative value than a study with similar results but a higher power.[98]

B. What Biases May Have Contributed to an Erroneous Association?

The second major reason for an invalid outcome in epidemiologic studies is systematic error or bias. Bias may arise in the design or conduct of a study, data collection, or data analysis. The meaning of scientific bias differs from conventional (and legal) usage, in which bias refers to a partisan point of view.[99] When scientists use the term *bias,* they refer to anything that results in a systematic (nonrandom) error in a study result and thereby compromises its validity. Two important categories of bias are selection bias (inappropriate methodology for selection of study subjects) and information bias (a flaw in measuring exposure or disease in the study groups).

Most epidemiologic studies have some degree of bias that may affect the outcome. If major bias is present, it may invalidate the study results. Finding the bias, however, can be difficult, if not impossible. In reviewing the validity of an epidemiologic study, the epidemiologist must identify potential biases and analyze the amount or kind of error that might have been induced by the bias. Often, the direction of error can be determined; depending on the specific type of bias, it may exaggerate the real association, dilute it, or even completely mask it.

1. Selection bias

Selection bias refers to the error in an observed association that results from the method of selection of cases and controls (in a case–control study) or exposed and unexposed individuals (in a cohort study).[100] The selection of an appropriate

---

98. *See also* David H. Kaye & David A. Freedman, Reference Guide on Statistics, Section IV.C.1, in this manual.

99. A Dictionary of Epidemiology 15 (John M. Last ed., 3d ed. 1995); Edmond A. Murphy, The Logic of Medicine 239–62 (1976).

100. Selection bias is defined as "[e]rror due to systematic differences in characteristics between those who are selected for study and those who are not." A Dictionary of Epidemiology, *supra* note 98, at 153. In *In re "Agent Orange" Product Liability Litigation,* 597 F. Supp. 740, 783 (E.D.N.Y. 1985), *aff'd,* 818 F.2d 145 (2d Cir. 1987), the court expressed concern about selection bias. The exposed cohort consisted of young, healthy men who served in Vietnam. Comparing the mortality rate of the exposed cohort and that of a control group made up of civilians might have resulted in error that was a result of selection bias. Failing to account for health status as an independent variable tends to understate any association between exposure and disease in studies in which the exposed cohort is healthier. *See also In re* Baycol Prods. Litig., 532 F. Supp. 2d 1029, 1043 (D. Minn. 2007) (upholding admissibility of testimony by expert witness who criticized study based on selection bias).

control group has been described as the Achilles' heel of a case–control study.[101] Ideally, controls should be drawn from the same population that produced the cases. Selecting control participants becomes problematic if the control participants are selected for reasons that are related to their having the exposure being studied. For example, a study of the effect of smoking on heart disease will suffer selection bias if subjects of the study are volunteers and the decision to volunteer is affected by both being a smoker and having a family history of heart disease. The association will be biased upward because of the additional disease among the exposed smokers caused by genetics.

Hospital-based studies, which are relatively common among researchers located in medical centers, illustrate the problem. Suppose an association is found between coffee drinking and coronary heart disease in a study using hospital patients as controls. The problem is that the hospitalized control group may include individuals who had been advised against drinking coffee for medical reasons, such as to prevent the aggravation of a peptic ulcer. In other words, the controls may become eligible for the study because of their medical condition, which is in turn related to their exposure status—their likelihood of avoiding coffee. If this is true, the amount of coffee drinking in the control group would understate the extent of coffee drinking expected in people who do not have the disease, and thus bias upwardly (i.e., exaggerate) any odds ratio observed.[102] Bias in hospital studies may also understate the true odds ratio when the exposures at issue led to the cases' hospitalizations and also contributed to the controls' chances of hospitalization.

Just as cases and controls in case-control studies should be selected independently of their exposure status, so the exposed and unexposed participants in cohort studies should be selected independently of their disease risk.[103] For example, if women with hysterectomies are overrepresented among exposed women in a cohort study of cervical cancer, this could overstate the association between the exposure and the disease.

A further source of selection bias occurs when those selected to participate decline to participate or drop out before the study is completed. Many studies have shown that individuals who participate in studies differ significantly from those who do not. If a significant portion of either study group declines to participate, the researcher should investigate whether those who declined are different from those who agreed. The researcher can compare relevant characteristics of those who

---

101. William B. Kannel & Thomas R. Dawber, *Coffee and Coronary Disease,* 289 New Eng. J. Med. 100 (1973) (editorial).

102. Hershel Jick et al., *Coffee and Myocardial Infarction,* 289 New Eng. J. Med. 63 (1973).

103. When unexposed controls may differ from the exposed cohort because exposure is associated with other risk (or protective factors), investigators can attempt to measure and adjust for those differences, as explained in Section IV.C.3, *infra. See also* Martha J. Radford & JoAnne M. Foody, *How Do Observational Studies Expand the Evidence Base for Therapy?* 286 JAMA 1228 (2001) (discussing the use of propensity analysis to adjust for potential confounding and selection biases that may occur from nonrandomization).

participate with those who do not to show the extent to which the two groups are comparable. Similarly, if a significant number of subjects drop out of a study before completion, the remaining subjects may not be representative of the original study populations. The researcher should examine whether that is the case.

The fact that a study may suffer from selection bias does not necessarily invalidate its results. A number of factors may suggest that a bias, if present, had only limited effect. If the association is particularly strong, for example, bias is less likely to account for all of it. In addition, a consistent association across different control groups suggests that possible biases applicable to a particular control group are not invalidating. Similarly, a dose–response relationship (*see* Section V.C, *infra*) found among multiple groups exposed to different doses of the agent would provide additional evidence that biases applicable to the exposed group are not a major problem.

2. Information bias

Information bias is a result of inaccurate information about either the disease or the exposure status of the study participants or a result of confounding. In a case-control study, potential information bias is an important consideration because the researcher depends on information from the past to determine exposure and disease and their temporal relationship.[104] In some situations, the researcher is required to interview the subjects about past exposures, thus relying on the sub-jects' memories. Research has shown that individuals with disease (cases) tend to recall past exposures more readily than individuals with no disease (controls);[105] this creates a potential for bias called recall bias.

For example, consider a case-control study conducted to examine the cause of congenital malformations. The epidemiologist is interested in whether the malfor-mations were caused by an infection during the mother's pregnancy.[106] A group of mothers of malformed infants (cases) and a group of mothers of infants with no

---

104. Information bias can be a problem in cohort studies as well. When exposure is determined retrospectively, there can be a variety of impediments to obtaining accurate information. Similarly, when disease status is determined retrospectively, bias is a concern. The determination that asbestos is a cause of mesothelioma was hampered by inaccurate death certificates that identified lung cancer rather than mesothelioma, a rare form of cancer, as the cause of death. *See* I.J. Selikoff et al., *Mortality Experi-ence of Insulation Workers in the United States and Canada,* 220 Ann. N.Y. Acad. Sci. 91, 110–11 (1979).

105. Steven S. Coughlin, *Recall Bias in Epidemiological Studies,* 43 J. Clinical Epidemiology 87 (1990).

106. *See* Brock v. Merrell Dow Pharms., Inc., 874 F.2d 307, 311–12 (5th Cir. 1989) (discus-sion of recall bias among women who bear children with birth defects). We note that the court was mistaken in its assertion that a confidence interval could correct for recall bias, or for any bias for that matter. Confidence intervals are a statistical device for analyzing error that may result from ran-dom sampling. Systematic errors (bias) in the design or data collection are not addressed by statistical methods, such as confidence intervals or statistical significance. *See* Green, *supra* note 47, at 667–68; Vincent M. Brannigan et al., *Risk, Statistical Inference, and the Law of Evidence: The Use of Epidemiological Data in Toxic Tort Cases,* 12 Risk Analysis 343, 344–45 (1992).

malformation (controls) are interviewed regarding infections during pregnancy. Mothers of children with malformations may recall an inconsequential fever or runny nose during pregnancy that readily would be forgotten by a mother who had a normal infant. Even if in reality the infection rate in mothers of malformed children is no different from the rate in mothers of normal children, the result in this study would be an apparently higher rate of infection in the mothers of the children with the malformations solely on the basis of recall differences between the two groups.[107] The issue of recall bias can sometimes be evaluated by finding an alternative source of data to validate the subject's response (e.g., blood test results from prenatal visits or medical records that document symptoms of infection).[108] Alternatively, the mothers' responses to questions about other exposures may shed light on the presence of a bias affecting the recall of the relevant exposures. Thus, if mothers of cases do not recall greater exposure than controls' mothers to pesticides, children with German measles, and so forth, then one can have greater confidence in their recall of illnesses.

Bias may also result from reliance on interviews with surrogates who are individuals other than the study subjects. This is often necessary when, for example, a subject (in a case-control study) has died of the disease under investigation or may be too ill to be interviewed.

There are many sources of information bias that affect the measure of exposure, including its intensity and duration. Exposure to the agent can be measured directly or indirectly.[109] Sometimes researchers use a biological marker as a direct measure of exposure to an agent—an alteration in tissue or body fluids that occurs as a result of an exposure and that can be detected in the laboratory. Biological markers, however, are only available for a small number of toxins and usually only reveal whether a person was exposed.[110] Biological markers rarely help determine the intensity or duration of exposure.[111]

---

107. Thus, in *Newman v. Motorola, Inc.,* 218 F. Supp. 2d 769, 778 (D. Md. 2002), the court considered a study of the effect of cell phone use on brain cancer and concluded that there was good reason to suspect that recall bias affected the results of the study, which found an association between cell phone use and cancers on the side of the head where the cell phone was used but no association between cell phone use and overall brain tumors.

108. Two researchers who used a case-control study to examine the association between congenital heart disease and the mother's use of drugs during pregnancy corroborated interview data with the mother's medical records. *See* Sally Zierler & Kenneth J. Rothman, *Congenital Heart Disease in Relation to Maternal Use of Bendectin and Other Drugs in Early Pregnancy,* 313 New Eng. J. Med. 347, 347–48 (1985).

109. *See In re* Paoli R.R. Yard PCB Litig., No. 86-2229, 1992 U.S. Dist LEXIS 18430, at ★9–★11 (E.D. Pa. Oct. 21, 1992) (discussing valid methods of determining exposure to chemicals).

110. *See* Gary E. Marchant, *Genetic Susceptibility and Biomarkers in Toxic Injury Litigation,* 41 Jurimetrics J. 67, 68, 73–74, 95–97 (2000) (explaining concept of biomarkers, how they might be used to provide evidence of exposure or dose, discussing cases in which biomarkers were invoked in an effort to prove exposure, and concluding, "biomarkers are likely to be increasingly relied on to demonstrate exposure").

111. There are different definitions of dose, but dose often refers to the intensity or magnitude of exposure multiplied by the time exposed. *See* Sparks v. Owens-Illinois, Inc., 38 Cal. Rptr. 2d 739,

Monitoring devices also can be used to measure exposure directly but often are not available for exposures that have occurred in the past. For past exposures, epidemiologists often use indirect measures of exposure, such as interviewing workers and reviewing employment records. Thus, all those employed to install asbestos insulation may be treated as having been exposed to asbestos during the period that they were employed. However, there may be a wide variation of exposure within any job, and these measures may have limited applicability to a given individual.[112] If the agent of interest is a drug, medical or hospital records can be used to determine past exposure. Thus, retrospective studies, which are often used for occupational or environmental investigations, entail measurements of exposure that are usually less accurate than prospective studies or followup studies, including ones in which a drug or medical intervention is the independent variable being measured.

---

742 (Ct. App. 1995). Other definitions of dose may be more appropriate in light of the biological mechanism of the disease.

For a discussion of the difficulties of determining dose from atomic fallout, see Allen v. United States, 588 F. Supp. 247, 425–26 (D. Utah 1984), *rev'd on other grounds,* 816 F.2d 1417 (10th Cir. 1987). The timing of exposure may also be critical, especially if the disease of interest is a birth defect. In *Smith v. Ortho Pharmaceutical Corp.,* 770 F. Supp. 1561, 1577 (N.D. Ga. 1991), the court criticized a study for its inadequate measure of exposure to spermicides. The researchers had defined exposure as receipt of a prescription for spermicide within 600 days of delivery, but this definition of exposure is too broad because environmental agents are likely to cause birth defects only during a narrow band of time.

A different, but related, problem often arises in court. Determining the plaintiff's exposure to the alleged toxic substance always involves a retrospective determination and may involve difficulties similar to those faced by an epidemiologist planning a study. Thus, in *John's Heating Service v. Lamb,* 46 P.3d 1024 (Alaska 2002), plaintiffs were exposed to carbon monoxide because of defendants' negligence with respect to a home furnace. The court observed: "[W]hile precise information concerning the exposure necessary to cause specific harm to humans and exact details pertaining to the plaintiff's exposure are beneficial, such evidence is not always available, or necessary, to demonstrate that a substance is toxic to humans given substantial exposure and need not invariably provide the basis for an expert's opinion on causation." *Id.* at 1035 (quoting Westberry v. Gislaved Gummi AB, 178 F.3d 257, 264 (4th Cir. 1999)); *see also* Alder v. Bayer Corp., AGFA Div., 61 P.3d 1068, 1086–88 (Utah 2002) (summarizing other decisions on the precision with which plaintiffs must establish the dosage to which they were exposed). *See generally* Restatement (Third) of Torts: Liability for Physical and Emotional Harm § 28 cmt. c(2) & rptrs. note (2010).

In asbestos litigation, a number of courts have adopted a requirement that the plaintiff demonstrate (1) regular use by an employer of the defendant's asbestos-containing product, (2) the plaintiff's proximity to that product, and (3) exposure over an extended period of time. *See, e.g.,* Lohrmann v. Pittsburgh Corning Corp., 782 F.2d 1156, 1162–64 (4th Cir. 1986); Gregg v. V-J Auto Parts, Inc., 943 A.2d 216, 226 (Pa. 2007).

112. Frequently, occupational epidemiologists employ study designs that consider all agents to which those who work in a particular occupation are exposed because they are trying to determine the hazards associated with that occupation. Isolating one of the agents for examination would be difficult if not impossible. These studies, then, present difficulties when employed in court in support of a claim by a plaintiff who was exposed to only one or fewer than all of the agents present at the worksite that was the subject of the study. *See, e.g.,* Knight v. Kirby Inland Marine Inc., 482 F.3d 347, 352–53 (5th Cir. 2007) (concluding that case-control studies of cancer that entailed exposure to a variety of organic solvents at job sites did not support claims of plaintiffs who claimed exposure to benzene caused their cancers).

The route (e.g., inhalation or absorption), duration, and intensity of exposure are important factors in assessing disease causation. Even with environmental monitoring, the dose measured in the environment generally is not the same as the dose that reaches internal target organs. If the researcher has calculated the internal dose of exposure, the scientific basis for this calculation should be examined for soundness.[113]

In assessing whether the data may reflect inaccurate information, one must assess whether the data were collected from objective and reliable sources. Medical records, government documents, employment records, death certificates, and interviews are examples of data sources that are used by epidemiologists to measure both exposure and disease status.[114] The accuracy of a particular source may affect the validity of a research finding. If different data sources are used to collect information about a study group, differences in the accuracy of those sources may affect the validity of the findings. For example, using employment records to gather information about exposure to narcotics probably would lead to inaccurate results, because employees tend to keep such information private. If the researcher uses an unreliable source of data, the study may not be useful.

The kinds of quality control procedures used may affect the accuracy of the data. For data collected by interview, quality control procedures should probe the reliability of the individual and whether the information is verified by other sources. For data collected and analyzed in the laboratory, quality control procedures should probe the validity and reliability of the laboratory test.

Information bias may also result from inaccurate measurement of disease status. The quality and sophistication of the diagnostic methods used to detect a disease should be assessed.[115] The proportion of subjects who were examined also should be questioned. If, for example, many of the subjects refused to be tested, the fact that the test used was of high quality would be of relatively little value.

---

113. *See also* Bernard D. Goldstein & Mary Sue Henifin, Reference Guide on Toxicology, Section I.D, in this manual.

114. Even these sources may produce unanticipated error. Identifying the causal connection between asbestos and mesothelioma, a rare form of cancer, was complicated and delayed because doctors who were unfamiliar with mesothelioma erroneously identified other causes of death in death certificates. *See* David E. Lilienfeld & Paul D. Gunderson, *The "Missing Cases" of Pleural Malignant Mesothelioma in Minnesota, 1979–81: Preliminary Report,* 101 Pub. Health Rep. 395, 397–98 (1986).

115. The hazards of adversarial review of epidemiologic studies to determine bias is highlighted by *O'Neill v. Novartis Consumer Health, Inc.,* 55 Cal. Rptr. 3d 551, 558–60 (Ct. App. 2007). Defendant's experts criticized a case–control study relied on by plaintiff on the ground that there was misclassification of exposure status among the cases. Plaintiff objected to this criticism because defendant's experts had only examined the cases for exposure misclassification, which would tend to exaggerate any association by providing an inaccurately inflated measure of exposure in the cases. The experts failed to examine whether there was misclassification in the controls, which, if it existed, would tend to incorrectly diminish any association.

The scientific validity of the research findings is influenced by the reliability of the diagnosis of disease or health status under study.[116] The disease must be one that is recognized and defined to enable accurate diagnoses.[117] Subjects' health status may be essential to the hypothesis under investigation. For example, a researcher interested in studying spontaneous abortion in the first trimester must determine that study subjects are pregnant. Diagnostic criteria that are accepted by the medical community should be used to make the diagnosis. If a diagnosis had been made at a time when home pregnancy kits were known to have a high rate of false-positive results (indicating pregnancy when the woman is not pregnant), the study will overestimate the number of spontaneous abortions.

Misclassification bias is a consequence of information bias in which, because of problems with the information available, individuals in the study may be misclassified with regard to exposure status or disease status. Bias due to exposure misclassification can be differential or nondifferential. In nondifferential misclassification, the inaccuracies in determining exposure are independent of disease status, or the inaccuracies in diagnoses are independent of exposure status—in other words, the data are crude, with a great deal of random error. This is a common problem. Generally, nondifferential misclassification bias leads to a shift in the odds ratio toward one, or, in other words, toward a finding of no effect. Thus, if the errors are nondifferential, it is generally misguided to criticize an apparent association between an exposure and disease on the ground that data were inaccurately classified. Instead, nondifferential misclassification generally underestimates the true size of the association.

Differential misclassification is systematic error in determining exposure in cases as compared with controls, or disease status in unexposed cohorts relative to exposed cohorts. In a case–control study this would occur, for example, if, in the

---

116. In *In re Swine Flu Immunization Products Liability Litigation,* 508 F. Supp. 897, 903 (D. Colo. 1981), *aff'd sub nom. Lima v. United States,* 708 F.2d 502 (10th Cir. 1983), the court critically evaluated a study relied on by an expert whose testimony was stricken. In that study, determination of whether a patient had Guillain-Barré syndrome was made by medical clerks, not physicians who were familiar with diagnostic criteria.

117. The difficulty of ill-defined diseases arose in some of the silicone gel breast implant cases. Thus, in *Grant v. Bristol-Myers Squibb,* 97 F. Supp. 2d 986 (D. Ariz. 2000), in the face of a substantial body of exonerative epidemiologic evidence, the female plaintiff alleged she suffered from an atypical systemic joint disease. The court concluded:

> As a whole, the Court finds that the evidence regarding systemic disease as proposed by Plaintiffs' experts is not scientifically valid and therefore will not assist the trier of fact. As for the atypical syndrome that is suggested, where experts propose that breast implants cause a disease but cannot specify the criteria for diagnosing the disease, it is incapable of epidemiologic testing. This renders the experts' methods insufficiently reliable to help the jury.

Id. at 992; *see also* Burton v. Wyeth-Ayerst Labs., 513 F. Supp. 2d 719, 722–24 (N.D. Tex. 2007) (parties disputed whether cardiology problem involved two separate diseases or only one; court concluded that all experts in the case reflected a view that there was but a single disease); *In re* Breast Implant Cases, 942 F. Supp. 958, 961 (E.D.N.Y. & S.D.N.Y. 1996).

process of anguishing over the possible causes of the disease, parents of ill children recalled more exposures to a particular agent than actually occurred, or if parents of the controls, for whom the issue was less emotionally charged, recalled fewer. This can also occur in a cohort study in which, for example, birth control users (the exposed cohort) are monitored more closely for potential side effects, leading to a higher rate of disease identification in that cohort than in the unexposed cohort. Depending on how the misclassification occurs, a differential bias can produce an error in either direction—the exaggeration or understatement of a true association.

3. Other conceptual problems

There are dozens of other potential biases that can occur in observational studies, which is an important reason why clinical studies (when ethical) are often preferable. Sometimes studies are limited by flawed definitions or premises. For example, if the researcher defines the disease of interest as all birth defects, rather than a specific birth defect, there should be a scientific basis to hypothesize that the effects of the agent being investigated could be so broad. If the effect is in fact more limited, the result of this conceptualization error could be to dilute or mask any real effect that the agent might have on a specific type of birth defect.[118]

Some biases go beyond errors in individual studies and affect the overall body of available evidence in a way that skews what appears to be the universe of evidence. Publication bias is the tendency for medical journals to prefer studies that find an effect.[119] If negative studies are never published, the published literature will be biased. Financial conflicts of interest by researchers and the source of funding of studies have been shown to have an effect on the outcomes of such studies.[120]

---

118. In *Brock v. Merrell Dow Pharmaceuticals, Inc.*, 874 F.2d 307, 312 (5th Cir. 1989), the court discussed a reanalysis of a study in which the effect was narrowed from all congenital malformations to limb reduction defects. The magnitude of the association changed by 50% when the effect was defined in this narrower fashion. *See* Rothman et al. *supra* note 61, at 144 ("Unwarranted assurances of a lack of any effect can easily emerge from studies in which a wide range of etiologically unrelated outcomes are grouped.").

119. Investigators may contribute to this effect by neglecting to submit negative studies for publication.

120. *See* Jerome P. Kassirer, On the Take: How Medicine's Complicity with Big Business Can Endanger Your Health 79–84 (2005); J.E. Bekelman et al., *Scope and Impact of Financial Conflicts of Interest in Biomedical Research: A Systematic Review,* 289 JAMA 454 (2003). Richard Smith, the editor in chief of the British Medical Journal, wrote on this subject:

> The major determinant of whether reviews of passive smoking concluded it was harmful was whether the authors had financial ties with tobacco manufacturers. In the disputed topic of whether third-generation contraceptive pills cause an increase in thromboembolic disease, studies funded by the pharmaceutical industry find that they don't and studies funded by public money find that they do.

Richard Smith, *Making Progress with Competing Interests,* 325 Brit. Med. J. 1375, 1376 (2002).

Examining a study for potential sources of bias is an important task that helps determine the accuracy of a study's conclusions. In addition, when a source of bias is identified, it may be possible to determine whether the error tended to exaggerate or understate the true association. Thus, bias may exist in a study that nevertheless has probative value.

Even if one concludes that the findings of a study are statistically stable and that biases have not created significant error, additional considerations remain. As repeatedly noted, an association does not necessarily mean a causal relationship exists. To make a judgment about causation, a knowledgeable expert[121] must consider the possibility of confounding factors. The expert must also evaluate several criteria to determine whether an inference of causation is appropriate.[122] These matters are discussed below.

C. Could a Confounding Factor Be Responsible for the Study Result?[123]

The third major reason for error in epidemiologic studies is confounding. Confounding occurs when another causal factor (the confounder) confuses the relationship between the agent of interest and outcome of interest.[124] (Confounding and selection bias (Section IV.B.1, *supra*) can, depending on terminology, overlap.) Thus, one instance of confounding is when a confounder is both a risk factor for the disease and a factor associated with the exposure of interest. For example, researchers may conduct a study that finds individuals with gray hair have a higher rate of death than those with hair of another color. Instead of hair color having an impact on death, the results might be explained by the confounding factor of age. If old age is associated differentially with the gray-haired group (those with gray hair tend to be older), old age may be responsible for the association found between hair color and death.[125] Researchers must separate the relationship between gray hair and risk of death from that of old age and risk of death. When researchers find an association between an agent and a disease, it is critical to determine whether the association is causal or the result of confounding.[126] Some

---

121. In a lawsuit, this would be done by an expert. In science, the effort is usually conducted by a panel of experts.

122. For an excellent example of the authors of a study analyzing whether an inference of causation is appropriate in a case–control study examining whether bromocriptine (Parlodel)—a lactation suppressant—causes seizures in postpartum women, see Kenneth J. Rothman et al., *Bromocriptine and Puerpal Seizures,* 1 Epidemiology 232, 236–38 (1990).

123. *See* Grassis v. Johns-Manville Corp., 591 A.2d 671, 675 (N.J. Super. Ct. App. Div. 1991) (discussing the possibility that confounders may lead to an erroneous inference of a causal relationship).

124. *See* Rothman et al., *supra* note 61, at 129.

125. This example is drawn from Kahn & Sempos, *supra* note 31, at 63.

126. Confounding can bias a study result by either exaggerating or diluting any true association. One example of a confounding factor that may result in a study's outcome understating an

epidemiologists classify confounding as a form of bias. However, confounding is a reality—that is, the observed association of a factor and a disease is actually the result of an association with a third, confounding factor.[127]

Confounding can be illustrated by a hypothetical prospective cohort study of the role of alcohol consumption and emphysema. The study is designed to investigate whether drinking alcohol is associated with emphysema. Participants are followed for a period of 20 years and the incidence of emphysema in the "exposed" (participants who consume more than 15 drinks per week) and the unexposed is compared. At the conclusion of the study, the relative risk of emphysema in the drinking group is found to be 2.0, an association that suggests a possible effect). But does this association reflect a true causal relationship or might it be the product of confounding?

One possibility for a confounding factor is smoking, a known causal risk factor for emphysema. If those who drink alcohol are more likely to be smokers than those who do not drink, then smoking may be responsible for some or all of the higher level of emphysema among those who do not drink.

A serious problem in observational studies such as this hypothetical study is that the individuals are not assigned randomly to the groups being compared.[128] As discussed above, randomization maximizes the possibility that exposures other than the one under study are evenly distributed between the exposed and the control cohorts.[129] In observational studies, by contrast, other forces, including self-selection, determine who is exposed to other (possibly causal) factors. The lack of randomization leads to the potential problem of confounding. Thus, for example, the exposed cohort might consist of those who are exposed at work to an agent suspected of being an industrial toxin. The members of this cohort may, however, differ from unexposed controls by residence, socioeconomic or health status, age, or other extraneous factors.[130] These other factors may be causing (or

---

association is vaccination. Thus, if a group exposed to an agent has a higher rate of vaccination for the disease under study than the unexposed group, the vaccination may reduce the rate of disease in the exposed group, thereby producing an association that is less than the true association without the confounding of vaccination.

127. *Schwab v. Philip Morris USA, Inc.*, 449 F. Supp. 2d 992, 1199–1200 (E.D.N.Y. 2006), *rev'd on other grounds,* 522 F.3d 215 (2d Cir. 2008), describes confounding that led to premature conclusions that low-tar cigarettes were safer than regular cigarettes. Smokers who chose to switch to low-tar cigarettes were different from other smokers in that they were more health conscious in other aspects of their lifestyles. Failure to account for that confounding—and measuring a healthy lifestyle is difficult even if it is identified as a potential confounder—biased the results of those studies.

128. Randomization attempts to ensure that the presence of a characteristic, such as coffee drinking, is governed by chance, as opposed to being determined by the presence of an underlying medical condition.

129. *See* Rothman et al., *supra* note 61, at 129; *see also supra* Section II.A.

130. *See, e.g., In re* "Agent Orange" Prod. Liab. Litig., 597 F. Supp. 740, 783 (E.D.N.Y. 1984) (discussing the problem of confounding that might result in a study of the effect of exposure to Agent Orange on Vietnam servicemen), *aff'd,* 818 F.2d 145 (2d Cir. 1987).

protecting against) the disease, but because of potential confounding, an apparent (yet false) association of the disease with exposure to the agent may appear. Confounders, like smoking in the alcohol drinking study, do not reflect an error made by the investigators; rather, they reflect the inherently "uncontrolled" nature of exposure designations in observational studies. When they can be identified, confounders should be taken into account. Unanticipated confounding factors that are suspected after data collection can sometimes be controlled during data analysis, if data have been gathered about them.

To evaluate whether smoking is a confounding factor, the researcher would stratify each of the exposed and control groups into smoking and nonsmoking subgroups to examine whether subjects' smoking status affects the study results. If the relationship between alcohol drinking and emphysema in the smoking subgroups is the same as that in the all-subjects group, smoking is not a confounding factor. If the subjects' smoking status affects the relationship between drinking and emphysema, then smoking is a confounder, for which adjustment is required. If the association between drinking and emphysema completely disappears when the subjects' smoking status is considered, then smoking is a confounder that fully accounts for the association with drinking observed. Table 4 reveals our hypothetical study's results, with smoking being a confounding factor, which, when accounted for, eliminates the association. Thus, in the full cohort, drinkers have twice the risk of emphysema compared with nondrinkers. When the relationship between drinking and emphysema is examined separately in smokers and in nonsmokers, the risk of emphysema in drinkers compared with nondrinkers is not elevated in smokers or in nonsmokers. This is because smokers are disproportionately drinkers and have a higher rate of emphysema than nonsmokers. Thus, the relationship between drinking and emphysema in the full cohort is distorted by failing to take into account the relationship between being a drinker and a smoker.

Even after accounting for the effect of smoking, there is always a risk that an undiscovered or unrecognized confounding factor may contribute to a study's findings, by either magnifying or reducing the observed association.[131] It is, however, necessary to keep that risk in perspective. Often the mere possibility of uncontrolled confounding is used to call into question the results of a study. This was certainly the strategy of some seeking, or unwittingly helping, to undermine the implications of the studies persuasively linking cigarette smoking to lung cancer. The critical question is whether it is plausible that the findings of a given study could indeed be due to unrecognized confounders.

In designing a study, researchers sometimes make assumptions that cannot be validated or evaluated empirically. Thus, researchers may assume that a missing potential confounder is not needed for the analysis or that a variable used was adequately classified. Researchers employ a sensitivity analysis to assess the effect of those assumptions should they be incorrect. Conducting a sensitivity analysis

---

131. Rothman et al., *supra* note 61, at 129; *see also supra* Section II.A.

Table 4. Hypothetical Emphysema Study Data[a]

Drinking Status	Total Cohort				Smokers				Nonsmokers			
	Total	Cases	Incidence	RR	Total	Cases	Incidence	RR	Total	Cases	Incidence	RR
Nondrinkers	471	16	0.034	1.0[b]	111	9	0.081	1.0[b]	360	7	0.019	1.0[b]
Drinkers	739	41	0.069	2.0	592	48	0.081	1.0	147	3	0.020	1.0

[a] The incidence of disease is not normally presented in an epidemiologic study, but we include it here to aid in comprehension of the ideas discussed in the text.
[b] RR = relative risk. The relative risk for each of the cohorts is determined based on reference to the risk among nondrinkers; that is, the incidence of disease among drinkers is compared with nondrinkers for each of the three cohorts separately.

entails repeating the analysis using different assumptions (e.g., alternative corrections for missing data or for classifying data) to see if the results are sensitive to the varying assumptions. Such analyses can show that the assumptions are not likely to affect the findings or that alternative explanations cannot be ruled out.[132]

1. What techniques can be used to prevent or limit confounding?

Choices in the design of a research project (e.g., methods for selecting the subjects) can prevent or limit confounding. In designing a study, the researcher must determine other risk factors for the disease under study. When a factor or factors, such as age, sex, or even smoking status, are risk factors and potential confounders in a study, investigators can limit the differential distribution of these factors in the study groups by selecting controls to "match" cases (or the exposed group) in terms of these variables. If the two groups are matched, for example, by age, then any association observed in the study cannot be due to age, the matched variable.[133]

Restricting the persons who are permitted as subjects in a study is another method to control for confounders. If age or sex is suspected as a confounder, then the subjects enrolled in a study can be limited to those of one sex and those who are within a specified age range. When there is no variance among subjects in a study with regard to a potential confounder, confounding as a result of that variable is eliminated.

2. What techniques can be used to identify confounding factors?

Once the study data are ready to be analyzed, the researcher must assess a range of factors that could influence risk. In the hypothetical study, the researcher would evaluate whether smoking is a confounding factor by comparing the incidence of emphysema in smoking alcohol drinkers with the incidence in nonsmoking alcohol drinkers. If the incidence is substantially the same, smoking is not a confounding factor (e.g., smoking does not distort the relationship between alcohol drinking and the development of emphysema). If the incidence is substantially different, but still exists in the nonsmoking group, then smoking is a confounder, but does not wholly account for the association with alcohol drinking. If the association disappears, then smoking is a confounder that fully accounts for the association observed.

---

132. Kenneth Rothman & Sander Greenland, Modern Epidemiology (2d ed. 1998).

133. Selecting a control population based on matched variables necessarily affects the representativeness of the selected controls and may affect how generalizable the study results are to the population at large. However, for a study to have merit, it must first be internally valid; that is, it must not be subject to unreasonable sources of bias or confounding. Only after a study has been shown to meet this standard does its universal applicability or generalizability to the population at large become an issue. When a study population is not representative of the general or target population, existing scientific knowledge may permit reasonable inferences about the study's broader applicability, or additional confirmatory studies of other populations may be necessary.

3. What techniques can be used to control for confounding factors?

A good study design will consider potential confounders and obtain data about them if possible. If researchers have good data on potential confounders, they can control for those confounders in the data analysis. There are several analytic approaches to account for the distorting effects of a confounder, including stratification or multivariate analysis. Stratification permits an investigator to evaluate the effect of a suspected confounder by subdividing the study groups based on a confounding factor. Thus, in Table 4, drinkers have been stratified based on whether they smoke (the suspected confounder). To take another example that entails a continuous rather than dichotomous potential confounder, let us say we are interested in the relationship between smoking and lung cancer but suspect that air pollution or urbanization may confound the relationship. Thus, an observed relationship between smoking and lung cancer could theoretically be due in part to pollution, if smoking were more common in polluted areas. We could address this issue by stratifying our data by degree of urbanization and look at the relationship between smoking and lung cancer in each urbanization stratum. Figure 5 shows actual age-adjusted lung cancer mortality rates per 100,000 person-years by urban or rural classification and smoking category.[134]

Figure 5: Age-adjusted lung cancer mortality rates per 100,000 person-years by urban or rural classification and smoking category.
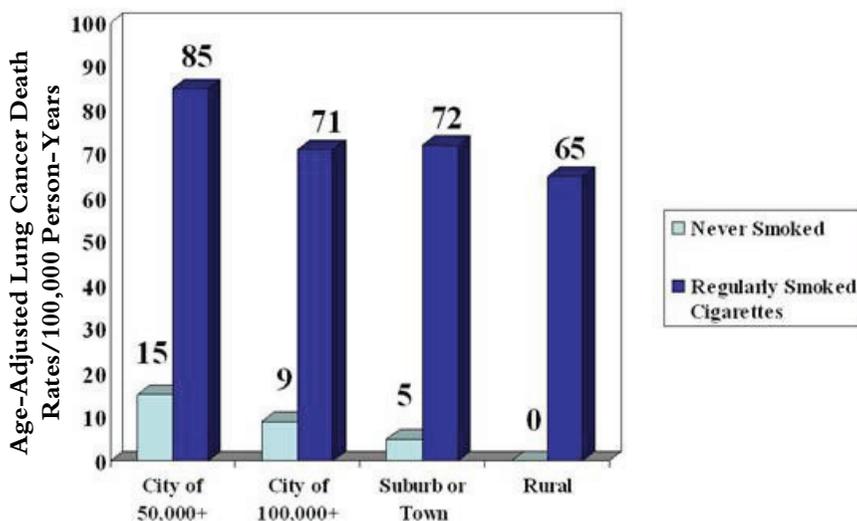
Source: Adapted from E. Cuyler Hammond & Daniel Horn, *Smoking and Death Rates—Report on Forty-Four Months of Follow-Up of 187,783 Men: II, Death Rates by Cause*, 166 JAMA 1294 (1958).

134. This example and Figure 4 are from Leon Gordis, Epidemiology 254 (4th ed. 2009).

For each degree of urbanization, lung cancer mortality rates in smokers are shown by the dark gray bars, and nonsmoker mortality rates are indicated by light gray bars. From these data we see that in every level (or stratum) of urbanization, lung cancer mortality is higher in smokers than in nonsmokers. Therefore, the observed association of smoking and lung cancer cannot be attributed to level of urbanization. By examining each stratum separately, we, in effect, hold urbanization constant, and still find much higher lung cancer mortality in smokers than in nonsmokers.

For each degree of urbanization, lung cancer mortality rates and smokers are shown by the dark-colored bars, and nonsmoker mortality rates are indicated by light-colored bars. For these data we see that in every level (or stratum) of urbanization, lung cancer mortality is higher in smokers than in nonsmokers. Therefore, the observed association of lung cancer cannot be attributed to level of urbanization. By examining each stratum separately, we are, in effect, holding urbanization constant, and we still find much higher lung cancer mortality in smokers than in nonsmokers.

Multivariate analysis controls for the confounding factor through mathematical modeling. Models are developed to describe the simultaneous effect of exposure and confounding factors on the increase in risk.[135]

Both of these methods allow for adjustment of the effect of confounders. They both modify an observed association to take into account the effect of risk factors that are not the subject of the study and that may distort the association between the exposure being studied and the disease outcomes. If the association between exposure and disease remains after the researcher completes the assessment and adjustment for confounding factors, the researcher must then assess whether an inference of causation is justified. This entails consideration of the Hill factors explained in Section V, *infra*.

V. General Causation: Is an Exposure a Cause of the Disease?

Once an association has been found between exposure to an agent and development of a disease, researchers consider whether the association reflects a true cause–effect relationship. When epidemiologists evaluate whether a cause–effect relationship exists between an agent and disease, they are using the term causation in a way similar to, but not identical to, the way that the familiar "but for," or sine qua non, test is used in law for cause in fact. "Conduct is a factual cause of

---

135. For a more complete discussion of multivariate analysis, see Daniel L. Rubinfeld, Reference Guide on Multiple Regression, in this manual.

[harm] when the harm would not have occurred absent the conduct."[136] This is equivalent to describing the conduct as a necessary link in a chain of events that results in the particular event.[137] Epidemiologists use causation to mean that an increase in the incidence of disease among the exposed subjects would not have occurred had they not been exposed to the agent.[138] Thus, exposure is a necessary condition for the increase in the incidence of disease among those exposed.[139] The relationship between the epidemiologic concept of cause and the legal question of whether exposure to an agent caused an individual's disease is addressed in Section VII.

As mentioned in Section I, epidemiology cannot prove causation; rather, causation is a judgment for epidemiologists and others interpreting the epidemiologic data.[140] Moreover, scientific determinations of causation are inherently tentative. The scientific enterprise must always remain open to reassessing the validity of past judgments as new evidence develops.

In assessing causation, researchers first look for alternative explanations for the association, such as bias or confounding factors, which are discussed in Section IV, *supra*. Once this process is completed, researchers consider how guidelines for inferring causation from an association apply to the available evidence. We emphasize that these guidelines are employed only *after* a study finds an association

---

136. Restatement (Third) of Torts: Liability for Physical and Emotional Harm § 26 (2010); *see also* Dan B. Dobbs, The Law of Torts § 168, at 409–11 (2000). When multiple causes are each operating and capable of causing an event, the but-for, or necessary-condition, concept for causation is problematic. This is the familiar "two-fires" scenario in which two independent fires simultaneously burn down a house and is sometimes referred to as overdetermined outcomes. Neither fire is a but-for, or necessary condition, for the destruction of the house, because either fire would have destroyed the house. *See* Restatement (Third) of Torts: Liability for Physical and Emotional Harm § 28 (2010). This two-fires situation is analogous to an individual being exposed to two agents, each of which is capable of causing the disease contracted by the individual. *See* Basko v. Sterling Drug, Inc., 416 F.2d 417 (2d Cir. 1969). A difference between the disease scenario and the fire scenario is that, in the former, one will have no more than a probabilistic assessment of whether each of the exposures would have caused the disease in the individual.

137. *See supra* note 7; *see also* Restatement (Third) of Torts: Liability for Physical and Emotional Harm § 26 cmt. c (2010) (employing a "causal set" model to explain multiple elements, each of which is required for an outcome).

138. "The imputed causal association is at the group level, and does not indicate the cause of disease in individual subjects." Bruce G. Charlton, *Attribution of Causation in Epidemiology: Chain or Mosaic?* 49 J. Clin. Epidemiology 105, 105 (1999).

139. *See* Rothman et al., *supra* note 61, at 8 ("We can define a cause of a specific disease event as an antecedent event, condition, or characteristic that was necessary for the occurrence of the disease at the moment it occurred, given that other conditions are fixed."); Allen v. United States, 588 F. Supp. 247, 405 (D. Utah 1984) (quoting a physician on the meaning of the statement that radiation causes cancer), *rev'd on other grounds,* 816 F.2d 1417 (10th Cir. 1987).

140. Restatement (Third) of Torts: Liability for Physical and Emotional Harm § 28 cmt. c (2010) ("[A]n evaluation of data and scientific evidence to determine whether an inference of causation is appropriate requires judgment and interpretation.").

to determine whether that association reflects a true causal relationship.[141] These guidelines consist of several key inquiries that assist researchers in making a judgment about causation.[142] Generally, researchers are conservative when it comes to assessing causal relationships, often calling for stronger evidence and more research before a conclusion of causation is drawn.[143]

The factors that guide epidemiologists in making judgments about causation (and there is no threshold number that must exist) are[144]

141. In a number of cases, experts attempted to use these guidelines to support the existence of causation in the absence of any epidemiologic studies finding an association. *See, e.g.,* Rains v. PPG Indus., Inc., 361 F. Supp. 2d 829, 836–37 (S.D. Ill. 2004) (explaining Hill criteria and proceeding to apply them even though there was no epidemiologic study that found an association); Soldo v. Sandoz Pharms. Corp., 244 F. Supp. 2d 434, 460–61 (W.D. Pa. 2003). There may be some logic to that effort, but it does not reflect accepted epidemiologic methodology. *See In re* Fosamax Prods. Liab. Litig., 645 F. Supp. 2d 164, 187–88 (S.D.N.Y. 2009); Dunn v. Sandoz Pharms. Corp., 275 F. Supp. 2d 672, 678–79 (M.D.N.C. 2003) ("The greater weight of authority supports Sandoz' assertion that [use of] the Bradford Hill criteria is a method for determining whether the results of an epidemiologic study can be said to demonstrate causation and not a method for testing an unproven hypothesis."); *Soldo*, 244 F. Supp. 2d at 514 (the Hill criteria "were developed as a mean[s] of interpreting an established association based on a body of epidemiologic research for the purpose of trying to judge whether the observed association reflects a causal relation between an exposure and disease." (quoting report of court-appointed expert)).

142. *See* Mervyn Susser, Causal Thinking in the Health Sciences: Concepts and Strategies in Epidemiology (1973); Gannon v. United States, 571 F. Supp. 2d 615, 624 (E.D. Pa. 2007) (quoting expert who testified that the Hill criteria are "'well-recognized' and widely used in the science community to assess general causation"); Chapin v. A & L Parts, Inc., 732 N.W.2d 578, 584 (Mich. Ct. App. 2007) (expert testified that Hill criteria are the most well-utilized method for determining if an association is causal).

143. Berry v. CSX Transp., Inc., 709 So. 2d 552, 568 n.12 (Fla. Dist. Ct. App. 1998) ("Almost all genres of research articles in the medical and behavioral sciences conclude their discussion with qualifying statements such as 'there is still much to be learned.' This is not, as might be assumed, an expression of ignorance, but rather an expression that all scientific fields are open-ended and can progress from their present state. . . ."); Hall v. Baxter Healthcare Corp., 947 F. Supp. 1387 app. B. at 1446–51 (D. Or. 1996) (report of Merwyn R. Greenlick, court-appointed epidemiologist). In *Cadarian v. Merrell Dow Pharmaceuticals, Inc.,* 745 F. Supp. 409 (E.D. Mich. 1989), the court refused to permit an expert to rely on a study that the authors had concluded should not be used to support an inference of causation in the absence of independent confirmatory studies. The court did not address the question whether the degree of certainty used by epidemiologists before making a conclusion of cause was consistent with the legal standard. *See* DeLuca v. Merrell Dow Pharms., Inc., 911 F.2d 941, 957 (3d Cir. 1990) (standard of proof for scientific community is not necessarily appropriate standard for expert opinion in civil litigation); Wells v. Ortho Pharm. Corp., 788 F.2d 741, 745 (11th Cir. 1986).

144. *See* Cook v. Rockwell Int'l Corp., 580 F. Supp. 2d 1071, 1098 (D. Colo. 2006) ("Defendants cite no authority, scientific or legal, that compliance with all, or even one, of these factors is required. . . . The scientific consensus is, in fact, to the contrary. It identifies Defendants' list of factors as some of the nine factors or lenses that guide epidemiologists in making judgments about causation. . . . These factors are not tests for determining the reliability of any study or the causal inferences drawn from it.").

1. Temporal relationship,
2. Strength of the association,
3. Dose–response relationship,
4. Replication of the findings,
5. Biological plausibility (coherence with existing knowledge),
6. Consideration of alternative explanations,
7. Cessation of exposure,
8. Specificity of the association, and
9. Consistency with other knowledge.

There is no formula or algorithm that can be used to assess whether a causal inference is appropriate based on these guidelines.[145] One or more factors may be absent even when a true causal relationship exists.[146] Similarly, the existence of some factors does not ensure that a causal relationship exists. Drawing causal inferences after finding an association and considering these factors requires judgment and searching analysis, based on biology, of why a factor or factors may be absent despite a causal relationship, and vice versa. Although the drawing of causal inferences is informed by scientific expertise, it is not a determination that is made by using an objective or algorithmic methodology.

These guidelines reflect criteria proposed by the U.S. Surgeon General in 1964[147] in assessing the relationship between smoking and lung cancer and expanded upon by Sir Austin Bradford Hill in 1965[148] and are often referred to as the Hill criteria or Hill factors.

145. *See* Douglas L. Weed, *Epidemiologic Evidence and Causal Inference,* 14 Hematology/Oncology Clinics N. Am. 797 (2000).

146. *See* Cook v. Rockwell Int'l Corp., 580 F. Supp. 2d 1071, 1098 (D. Colo. 2006) (rejecting argument that plaintiff failed to provide sufficient evidence of causation based on failing to meet four of the Hill factors).

147. Public Health Serv., U.S. Dep't of Health, Educ., & Welfare, Smoking and Health: Report of the Advisory Committee to the Surgeon General (1964); *see also* Centers for Disease Control and Prevention, U.S. Dep't of Health & Human Servs., The Health Consequences of Smoking: A Report of the Surgeon General (2004).

148. *See* Austin Bradford Hill, *The Environment and Disease: Association or Causation?* 58 Proc. Royal Soc'y Med. 295 (1965) (Hill acknowledged that his factors could only serve to assist in the inferential process: "None of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a *sine qua non*."). For discussion of these criteria and their respective strengths in informing a causal inference, see Gordis, *supra* note 32, at 236–39; David E. Lilienfeld & Paul D. Stolley, Foundations of Epidemiology 263–66 (3d ed. 1994); Weed, *supra* note 144.

A. Is There a Temporal Relationship?

A temporal, or chronological, relationship must exist for causation to exist. If an exposure causes disease, the exposure must occur before the disease develops.[149] If the exposure occurs after the disease develops, it cannot have caused the disease. Although temporal relationship is often listed as one of many factors in assessing whether an inference of causation is justified, this aspect of a temporal relationship is a necessary factor: Without exposure before the disease, causation cannot exist.[150]

With regard to specific causation, a subject dealt with in detail in Section VII, *infra*, there may be circumstances in which a temporal relationship supports the existence of a causal relationship. If the latency period between exposure and outcome is known,[151] then exposure consistent with that information may lend credence to a causal relationship. This is particularly true when the latency period is short and competing causes are known and can be ruled out. Thus, if an individual suffers an acute respiratory response shortly after exposure to a suspected agent and other causes of that respiratory problem are known and can be ruled out, the temporal relationship involved supports the conclusion that a causal relationship exists.[152] Similarly, exposure outside a known latency period constitutes evidence, perhaps conclusive evidence, against the existence of causation.[153] On the other hand, when latency periods are lengthy, variable, or not known and a

149. *See* Carroll v. Litton Sys., Inc., No. B-C-88-253, 1990 U.S. Dist. LEXIS 16833, at *29 (W.D.N.C. 1990) ("[I]t is essential for . . . [the plaintiffs' medical experts opining on causation] to know that exposure preceded plaintiffs' alleged symptoms in order for the exposure to be considered as a possible cause of those symptoms. . . .").

150. Exposure during the disease initiation process may cause the disease to be more severe than it otherwise would have been without the additional dose.

151. When the latency period is known—or is known to be limited to a specific range of time— as is the case with the adverse effects of some vaccines, the time frame from exposure to manifestation of disease can be critical to determining causation.

152. For courts that have relied on temporal relationships of the sort described, see Bonner v. ISP Technologies, Inc., 259 F.3d 924, 930–31 (8th Cir. 2001) (giving more credence to the expert's opinion on causation for acute response based on temporal relationship than for chronic disease that plaintiff also developed); Heller v. Shaw Industries, Inc. 167 F.3d 146 (3d Cir. 1999); Westberry v. Gislaved Gummi AB, 178 F.3d 257 (4th Cir. 1999); Zuchowicz v. United States, 140 F.3d 381 (2d Cir. 1998); Creanga v. Jardal, 886 A.2d 633, 641 (N.J. 2005); Alder v. Bayer Corp., AGFA Div., 61 P.3d 1068, 1090 (Utah 2002) ("If a bicyclist falls and breaks his arm, causation is assumed without argument because of the temporal relationship between the accident and the injury [and, the court might have added, the absence of any plausible competing causes that might instead be responsible for the broken arm].").

153. *See In re* Phenylpropanolamine (PPA) Prods. Liab. Litig., 289 F. Supp. 2d 1230, 1238 (W.D. Wash. 2003) (determining expert testimony on causation for plaintiffs whose exposure was beyond known latency period was inadmissible).

substantial proportion of the disease is due to unknown causes, temporal relationship provides little beyond satisfying the requirement that cause precede effect.[154]

B. *How Strong Is the Association Between the Exposure and Disease?*[155]

The relative risk is one of the cornerstones for causal inferences.[156] Relative risk measures the strength of the association. The higher the relative risk, the greater the likelihood that the relationship is causal.[157] For cigarette smoking, for example, the estimated relative risk for lung cancer is very high, about 10.[158] That is, the risk of lung cancer in smokers is approximately 10 times the risk in nonsmokers.

A relative risk of 10, as seen with smoking and lung cancer, is so high that it is extremely difficult to imagine any bias or confounding factor that might account for it. The higher the relative risk, the stronger the association and the lower the chance that the effect is spurious. Although lower relative risks can reflect causality, the epidemiologist will scrutinize such associations more closely because there is a greater chance that they are the result of uncontrolled confounding or biases.

154. These distinctions provide a framework for distinguishing between cases that are largely dismissive of temporal relationships as supporting causation and others that find it of significant persuasiveness. *Compare* cases cited in note 151, *supra*, *with* Moore v. Ashland Chem. Inc., 151 F.3d 269, 278 (5th Cir. 1998) (giving little weight to temporal relationship in a case in which there were several plausible competing causes that may have been responsible for the plaintiff's disease), *and* Glastetter v. Novartis Pharms. Corp., 252 F.3d 986, 990 (8th Cir. 2001) (giving little weight to temporal relationship in case studies involving drug and stroke).

155. Assuming that an association is determined to be causal, the strength of the association plays an important role legally in determining the specific causation question—whether the agent caused an individual plaintiff's injury. *See infra* Section VII.

156. *See supra* Section III.A.

157. *See* Miller v. Pfizer, Inc., 196 F. Supp. 2d 1062, 1079 (D. Kan. 2002) (citing this reference guide); Landrigan v. Celotex Corp., 605 A.2d 1079, 1085 (N.J. 1992). The use of the strength of the association as a factor does not reflect a belief that weaker effects occur less frequently than stronger effects. *See* Green, *supra* note 47, at 652–53 n.39. Indeed, the apparent strength of a given agent is dependent on the prevalence of the other necessary elements that must occur with the agent to produce the disease, rather than on some inherent characteristic of the agent itself. *See* Rothman et al., *supra* note 61, at 9–11.

158. *See* Doll & Hill, *supra* note 6. The relative risk of lung cancer from smoking is a function of intensity and duration of dose (and perhaps other factors). *See* Karen Leffondré et al., *Modeling Smoking History: A Comparison of Different Approaches,* 156 Am. J. Epidemiology 813 (2002). The relative risk provided in the text is based on a specified magnitude of cigarette exposure.

C. Is There a Dose–Response Relationship?

A dose–response relationship means that the greater the exposure, the greater the risk of disease. Generally, higher exposures should increase the incidence (or severity) of disease.[159] However, some causal agents do not exhibit a dose–response relationship when, for example, there is a threshold phenomenon (i.e., an exposure may not cause disease until the exposure exceeds a certain dose).[160] Thus, a dose–response relationship is strong, but not essential, evidence that the relationship between an agent and disease is causal.[161]

159. *See* Newman v. Motorola, Inc., 218 F. Supp. 2d 769, 778 (D. Md. 2002) (recognizing importance of dose–response relationship in assessing causation).

160. The question whether there is a no-effect threshold dose is a controversial one in a variety of toxic substances areas. *See, e.g.,* Irving J. Selikoff, Disability Compensation for Asbestos-Associated Disease in the United States: Report to the U.S. Department of Labor 181–220 (1981); Paul Kotin, *Dose–Response Relationships and Threshold Concepts,* 271 Ann. N.Y. Acad. Sci. 22 (1976); K. Robock, *Based on Available Data, Can We Project an Acceptable Standard for Industrial Use of Asbestos? Absolutely,* 330 Ann. N.Y. Acad. Sci. 205 (1979); Ferebee v. Chevron Chem. Co., 736 F.2d 1529, 1536 (D.C. Cir. 1984) (dose–response relationship for low doses is "one of the most sharply contested questions currently being debated in the medical community"); *In re* TMI Litig. Consol. Proc., 927 F. Supp. 834, 844–45 (M.D. Pa. 1996) (discussing low-dose extrapolation and no-dose effects for radiation exposure).

Moreover, good evidence to support or refute the threshold-dose hypothesis is exceedingly unlikely because of the inability of epidemiology or animal toxicology to ascertain very small effects. *Cf*. Arnold L. Brown, *The Meaning of Risk Assessment,* 37 Oncology 302, 303 (1980). Even the shape of the dose–response curve—whether linear or curvilinear, and if the latter, the shape of the curve—is a matter of hypothesis and speculation. *See* Allen v. United States, 588 F. Supp. 247, 419–24 (D. Utah 1984), *rev'd on other grounds,* 816 F.2d 1417 (10th Cir. 1987); *In re* Bextra & Celebrex Mktg. Sales Practices & Prod. Liab. Litig., 524 F. Supp. 2d 1166, 1180 (N.D. Cal. 2007) (criticizing expert for "primitive" extrapolation of risk based on assumption of linear relationship of risk to dose); Troyen A. Brennan & Robert F. Carter, *Legal and Scientific Probability of Causation for Cancer and Other Environmental Disease in Individuals,* 10 J. Health Pol'y & L. 33, 43–44 (1985).

The idea that the "dose makes the poison" is a central tenet of toxicology and attributed to Paracelsus, in the sixteenth century. *See* Bernard D. Goldstein & Mary Sue Henifin, Reference Guide on Toxicology, Section I.A, in this manual. It does not mean that any agent is capable of causing any disease if an individual is exposed to a sufficient dose. Agents tend to have specific effects, *see infra* Section V.H., and this dictum reflects only the idea that there is a safe dose below which an agent does not cause any toxic effect. *See* Michael A Gallo, *History and Scope of Toxicology,* in Casarett and Doull's Toxicology: The Basic Science of Poisons 1, 4–5 (Curtis D. Klaassen ed., 7th ed. 2008). For a case in which a party made such a mistaken interpretation of Paracelsus, see Alder v. Bayer Corp., AGFA Div., 61 P.3d 1068, 1088 (Utah 2002). Paracelsus was also responsible for the initial articulation of the specificity tenet. *See infra* Section V.H.

161. Evidence of a dose–response relationship as bearing on whether an inference of general causation is justified is analytically distinct from determining whether evidence of the dose to which a plaintiff was exposed is required in order to establish specific causation. On the latter matter, see *infra* Section VII; Restatement (Third) of Torts: Liability for Physical and Emotional Harm § 28 cmt. c(2) & rptrs. note (2010).

D. Have the Results Been Replicated?

Rarely, if ever, does a single study persuasively demonstrate a cause–effect relationship.[162] It is important that a study be replicated in different populations and by different investigators before a causal relationship is accepted by epidemiologists and other scientists.[163]

The need to replicate research findings permeates most fields of science. In epidemiology, research findings often are replicated in different populations.[164] Consistency in these findings is an important factor in making a judgment about causation. Different studies that examine the same exposure–disease relationship generally should yield similar results. Although inconsistent results do not necessarily rule out a causal nexus, any inconsistencies signal a need to explore whether different results can be reconciled with causality.

E. Is the Association Biologically Plausible (Consistent with Existing Knowledge)?[165]

Biological plausibility is not an easy criterion to use and depends upon existing knowledge about the mechanisms by which the disease develops. When biological plausibility exists, it lends credence to an inference of causality. For example, the conclusion that high cholesterol is a cause of coronary heart disease is plausible because cholesterol is found in atherosclerotic plaques. However, observations have been made in epidemiologic studies that were not biologically plausible at the time but subsequently were shown to be correct.[166] When an observation is inconsistent with current biological knowledge, it should not be discarded, but

---

162. In *Kehm v. Procter & Gamble Co.,* 580 F. Supp. 890, 901 (N.D. Iowa 1982), *aff'd,* 724 F.2d 613 (8th Cir. 1983), the court remarked on the persuasive power of multiple independent studies, each of which reached the same finding of an association between toxic shock syndrome and tampon use.

163. This may not be the legal standard, however. *Cf.* Smith v. Wyeth-Ayerst Labs. Co., 278 F. Supp. 2d 684, 710 n.55 (W.D.N.C. 2003) (observing that replication is difficult to establish when there is only one study that has been performed at the time of trial).

164. *See* Cadarian v. Merrell Dow Pharms., Inc., 745 F. Supp. 409, 412 (E.D. Mich. 1989) (holding a study on Bendectin insufficient to support an expert's opinion, because "the study's authors themselves concluded that the results could not be interpreted without independent confirmatory evidence").

165. A number of courts have adverted to this criterion in the course of their discussions of causation in toxic substances cases. *E.g., In re* Phenylpropanolamine (PPA) Prods. Liab. Litig., 289 F. Supp. 2d 1230, 1247–48 (W.D. Wash. 2003); Cook v. United States, 545 F. Supp. 306, 314–15 (N.D. Cal. 1982) (discussing biological implausibility of a two-peak increase of disease when plotted against time); Landrigan v. Celotex Corp., 605 A.2d 1079, 1085–86 (N.J. 1992) (discussing the existence vel non of biological plausibility); *see also* Bernard D. Goldstein & Mary Sue Henifin, Reference Guide on Toxicology, Section III.E, in this manual.

166. *See In re* Rezulin Prods. Liab. Litig., 369 F. Supp. 2d 398, 405 (S.D.N.Y. 2005); *In re* Phenylpropanolamine (PPA) Prods. Liab. Litig., 289 F. Supp. 2d 1230, 1247 (W.D. Wash. 2003).

the observation should be confirmed before significance is attached to it. The saliency of this factor varies depending on the extent of scientific knowledge about the cellular and subcellular mechanisms through which the disease process works. The mechanisms of some diseases are understood quite well based on the available evidence, including from toxicologic research, whereas other mechanism explanations are merely hypothesized—although hypotheses are sometimes accepted under this factor.[167]

F. Have Alternative Explanations Been Considered?

The importance of considering the possibility of bias and confounding and ruling out the possibilities is discussed above.[168]

G. What Is the Effect of Ceasing Exposure?

If an agent is a cause of a disease, then one would expect that cessation of exposure to that agent ordinarily would reduce the risk of the disease. This has been the case, for example, with cigarette smoking and lung cancer. In many situations, however, relevant data are simply not available regarding the possible effects of ending the exposure. But when such data are available and eliminating exposure reduces the incidence of disease, this factor strongly supports a causal relationship.

H. Does the Association Exhibit Specificity?

An association exhibits specificity if the exposure is associated only with a single disease or type of disease.[169] The vast majority of agents do not cause a wide vari-

---

167. *See* Douglas L. Weed & Stephen D. Hursting, *Biologic Plausibility in Causal Inference: Current Methods and Practice,* 147 Am. J. Epidemiology 415 (1998) (examining use of this criterion in contemporary epidemiologic research and distinguishing between alternative explanations of what constitutes biological plausibility, ranging from mere hypotheses to "sufficient evidence to show how the factor influences a known disease mechanism").

168. *See supra* Sections IV.B–C.

169. This criterion reflects the fact that although an agent causes one disease, it does not necessarily cause other diseases. *See, e.g.,* Nelson v. Am. Sterilizer Co., 566 N.W.2d 671, 676–77 (Mich. Ct. App. 1997) (affirming dismissal of plaintiff's claims that chemical exposure caused her liver disorder, but recognizing that evidence supported claims for neuropathy and other illnesses); Sanderson v. Int'l Flavors & Fragrances, Inc., 950 F. Supp. 981, 996–98 (C.D. Cal. 1996); *see also* Taylor v. Airco, Inc., 494 F. Supp. 2d 21, 27 (D. Mass. 2007) (holding that plaintiff's expert could testify to causal relationship between vinyl chloride and one type of liver cancer for which there was only modest support given strong causal evidence for vinyl chloride and another type of liver cancer).

When a party claims that evidence of a causal relationship between an agent and one disease is relevant to whether the agent caused another disease, courts have required the party to show that

ety of effects. For example, asbestos causes mesothelioma and lung cancer and may cause one or two other cancers, but there is no evidence that it causes any other types of cancers. Thus, a study that finds that an agent is associated with many different diseases should be examined skeptically. Nevertheless, there may be causal relationships in which this guideline is not satisfied. Cigarette manufacturers have long claimed that because cigarettes have been linked to lung cancer, emphysema, bladder cancer, heart disease, pancreatic cancer, and other conditions, there is no specificity and the relationships are not causal. There is, however, at least one good reason why inferences about the health consequences of tobacco do not require specificity: Because tobacco and cigarette smoke are not in fact single agents but consist of numerous harmful agents, smoking represents exposure to multiple agents, with multiple possible effects. Thus, whereas evidence of specificity may strengthen the case for causation, lack of specificity does not necessarily undermine it where there is a good biological explanation for its absence.

I. Are the Findings Consistent with Other Relevant Knowledge?

In addressing the causal relationship of lung cancer to cigarette smoking, researchers examined trends over time for lung cancer and for cigarette sales in the United States. A marked increase in lung cancer death rates in men was observed, which appeared to follow the increase in sales of cigarettes. Had the increase in lung cancer deaths followed a decrease in cigarette sales, it might have given researchers pause. It would not have precluded a causal inference, but the inconsistency of the trends in cigarette sales and lung cancer mortality would have had to be explained.

VI. What Methods Exist for Combining the Results of Multiple Studies?

Not infrequently, the scientific record may include a number of epidemiologic studies whose findings differ. These may be studies in which one shows an association and the other does not, or studies that report associations, but of different

---

the mechanisms involved in development of the disease are similar. Thus, in *Austin v. Kerr-McGee Refining Corp.*, 25 S.W.3d 280 (Tex. App. 2000), the plaintiff suffered from a specific form of chronic leukemia. Studies demonstrated a causal relationship between benzene and all leukemias, but there was a paucity of evidence on the relationship between benzene and the specific form of leukemia from which plaintiff suffered. The court required that plaintiff's expert demonstrate the similarity of the biological mechanism among leukemias as a condition for the admissibility of his causation testimony, a requirement the court concluded had not been satisfied. *Accord In re* Bextra & Celebrex Mktg. Sales Practices & Prod. Liab. Litig., 524 F. Supp. 2d 1166, 1183 (N.D. Cal. 2007); Magistrini v. One Hour Martinizing Dry Cleaning, 180 F. Supp. 2d 584, 603 (D.N.J. 2002).

magnitude.[170] In view of the fact that studies may disagree and that often many of the studies are small and lack the statistical power needed for definitive conclusions, the technique of meta-analysis was developed, initially for clinical trials.[171] Meta-analysis is a method of pooling study results to arrive at a single figure to represent the totality of the studies reviewed.[172] It is a way of systematizing the time-honored approach of reviewing the literature, which is characteristic of science, and placing it in a standardized framework with quantitative methods for estimating risk. In a meta-analysis, studies are given different weights in proportion to the sizes of their study populations and other characteristics.[173]

Meta-analysis is most appropriate when used in pooling randomized experimental trials, because the studies included in the meta-analysis share the most significant methodological characteristics, in particular, use of randomized assignment of subjects to different exposure groups. However, often one is confronted with nonrandomized observational studies of the effects of possible toxic substances or agents. A method for summarizing such studies is greatly needed, but when meta-analysis is applied to observational studies—either case-control or cohort—it becomes more controversial.[174] The reason for this is that often methodological differences among studies are much more pronounced than they are in randomized trials. Hence, the justification for pooling the results and deriving a single estimate of risk, for example, is problematic.[175]

---

170. *See, e.g.,* Zandi v. Wyeth a/k/a Wyeth, Inc., No. 27-CV-06-6744, 2007 WL 3224242 (Minn. Dist. Ct. Oct. 15, 2007) (plaintiff's expert cited 40 studies in support of a causal relationship between hormone therapy and breast cancer; many studies found different magnitudes of increased risk).

171. *See In re* Paoli R.R. Yard PCB Litig., 916 F.2d 829, 856 (3d Cir. 1990), *cert. denied,* 499 U.S. 961 (1991); Hines v. Consol. Rail Corp., 926 F.2d 262, 273 (3d Cir. 1991); Allen v. Int'l Bus. Mach. Corp., No. 94-264-LON, 1997 U.S. Dist. LEXIS 8016, at ⋆71–⋆74 (meta-analysis of observational studies is a controversial subject among epidemiologists). Thus, contrary to the suggestion by at least one court, multiple studies with small numbers of subjects may be pooled to reduce the possibility of sampling error. *See In re* Joint E. & S. Dist. Asbestos Litig., 827 F. Supp. 1014, 1042 (S.D.N.Y. 1993) ("[N]o matter how many studies yield a positive but statistically insignificant SMR for colorectal cancer, the results remain statistically insignificant. Just as adding a series of zeros together yields yet another zero as the product, adding a series of positive but statistically insignificant SMRs together does not produce a statistically significant pattern."), rev'd, 52 F.3d 1124 (2d Cir. 1995); *see also supra* note 76.

172. For a nontechnical explanation of meta-analysis, along with case studies of a variety of scientific areas in which it has been employed, see Morton Hunt, How Science Takes Stock: The Story of Meta-Analysis (1997).

173. Petitti, *supra* note 88.

174. *See* Donna F. Stroup et al., *Meta-analysis of Observational Studies in Epidemiology: A Proposal for Reporting,* 283 JAMA 2008, 2009 (2000); Jesse A. Berlin & Carin J. Kim, *The Use of Meta-Analysis in Pharmacoepidemiology, in* Pharmacoepidemiology 681, 683–84 (Brian L. Strom ed., 4th ed. 2005).

175. On rare occasions, meta-analyses of both clinical and observational studies are available. *See, e.g., In re* Bextra & Celebrex Mktg. Sales Practices & Prod. Liab. Litig., 524 F. Supp. 2d 1166, 1175 (N.D. Cal. 2007) (referring to clinical and observational meta-analyses of low dose of a drug; both analyses failed to find any effect).

A number of problems and issues arise in meta-analysis. Should only published papers be included in the meta-analysis, or should any available studies be used, even if they have not been peer reviewed? Can the results of the meta-analysis itself be reproduced by other analysts? When there are several meta-analyses of a given relationship, why do the results of different meta-analyses often disagree? The appeal of a meta-analysis is that it generates a single estimate of risk (along with an associated confidence interval), but this strength can also be a weakness, and may lead to a false sense of security regarding the certainty of the estimate. A key issue is the matter of heterogeneity of results among the studies being summarized. If there is more variance among study results than one would expect by chance, this creates further uncertainty about the summary measure from the meta-analysis. Such differences can arise from variations in study quality, or in study populations or in study designs. Such differences in results make it harder to trust a single estimate of effect; the reasons for such differences need at least to be acknowledged and, if possible, explained.[176] People often tend to have an inordinate belief in the validity of the findings when a single number is attached to them, and many of the difficulties that may arise in conducting a meta-analysis, especially of observational studies such as epidemiologic ones, may consequently be overlooked.[177]

VII. What Role Does Epidemiology Play in Proving Specific Causation?

Epidemiology is concerned with the incidence of disease in populations, and epidemiologic studies do not address the question of the cause of an individual's disease.[178] This question, often referred to as specific causation, is beyond the

---

176. *See* Stroup et al., *supra* note 173 (recommending methodology for meta-analysis of observational studies).

177. Much has been written about meta-analysis recently and some experts consider the problems of meta-analysis to outweigh the benefits at the present time. For example, John Bailar has observed:

> [P]roblems have been so frequent and so deep, and overstatements of the strength of conclusions so extreme, that one might well conclude there is something seriously and fundamentally wrong with the method. For the present . . . I still prefer the thoughtful, old-fashioned review of the literature by a knowledgeable expert who explains and defends the judgments that are presented. We have not yet reached a stage where these judgments can be passed on, even in part, to a formalized process such as meta-analysis.

John C. Bailar III, *Assessing Assessments,* 277 Science 528, 529 (1997) (reviewing Morton Hunt, How Science Takes Stock (1997)); *see also Point/Counterpoint: Meta-analysis of Observational Studies,* 140 Am. J. Epidemiology 770 (1994).

178. *See* DeLuca v. Merrell Dow Pharms., Inc., 911 F.2d 941, 945 & n.6 (3d Cir. 1990) ("Epidemiological studies do not provide direct evidence that a particular plaintiff was injured by exposure to a substance."); *In re* Viagra Prods. Liab. Litig., 572 F. Supp. 2d 1071, 1078 (D. Minn. 2008) ("Epi-

domain of the science of epidemiology. Epidemiology has its limits at the point where an inference is made that the relationship between an agent and a disease is causal (general causation) and where the magnitude of excess risk attributed to the agent has been determined; that is, epidemiologists investigate whether an agent can cause a disease, not whether an agent did cause a specific plaintiff's disease.[179]

Nevertheless, the specific causation issue is a necessary legal element in a toxic substance case. The plaintiff must establish not only that the defendant's agent is capable of causing disease, but also that it did cause the plaintiff's disease. Thus, numerous cases have confronted the legal question of what is acceptable proof of specific causation and the role that epidemiologic evidence plays in answering that question.[180] This question is not a question that is addressed by epidemiology.[181] Rather, it is a legal question with which numerous courts

<hr/>

demiology focuses on the question of general causation (i.e., is the agent capable of causing disease?) rather than that of specific causation (i.e., did it cause a disease in a particular individual?)" (quoting the second edition of this reference guide)); *In re* Asbestos Litig,, 900 A.2d 120, 133 (Del. Super. Ct. 2006); Michael Dore, *A Commentary on the Use of Epidemiological Evidence in Demonstrating Cause-in-Fact,* 7 Harv. Envtl. L. Rev. 429, 436 (1983).

There are some diseases that do not occur without exposure to a given toxic agent. This is the same as saying that the toxic agent is a necessary cause for the disease, and the disease is sometimes referred to as a signature disease (also, the agent is pathognomonic), because the existence of the disease necessarily implies the causal role of the agent. *See* Kenneth S. Abraham & Richard A. Merrill, *Scientific Uncertainty in the Courts,* Issues Sci. & Tech. 93, 101 (1986). Asbestosis is a signature disease for asbestos, and vaginal adenocarcinoma (in young adult women) is a signature disease for in utero DES exposure.

179. *Cf. In re* "Agent Orange" Prod. Liab. Litig., 597 F. Supp. 740, 780 (E.D.N.Y. 1984) (Agent Orange allegedly caused a wide variety of diseases in Vietnam veterans and their offspring), *aff'd,* 818 F.2d 145 (2d Cir. 1987).

180. In many instances, causation can be established without epidemiologic evidence. When the mechanism of causation is well understood, the causal relationship is well established, or the timing between cause and effect is close, scientific evidence of causation may not be required. This is frequently the situation when the plaintiff suffers traumatic injury rather than disease. This section addresses only those situations in which causation is not evident, and scientific evidence is required.

181. Nevertheless, an epidemiologist may be helpful to the factfinder in answering this question. Some courts have permitted epidemiologists (or those who use epidemiologic methods) to testify about specific causation. *See* Ambrosini v. Labarraque, 101 F.3d 129, 137–41 (D.C. Cir. 1996); Zuchowicz v. United States, 870 F. Supp. 15 (D. Conn. 1994); Landrigan v. Celotex Corp., 605 A.2d 1079, 1088–89 (N.J. 1992). In general, courts seem more concerned with the basis of an expert's opinion than with whether the expert is an epidemiologist or clinical physician. *See* Porter v. Whitehall, 9 F.3d 607, 614 (7th Cir. 1992) ("curb side" opinion from clinician not admissible); Burton v. R.J. Reynolds Tobacco Co., 181 F. Supp. 2d 1256, 1266–67 (D. Kan. 2002) (vascular surgeon permitted to testify to general causation over objection based on fact he was not an epidemiologist); Wade-Greaux v. Whitehall Labs., 874 F. Supp. 1441, 1469–72 (D.V.I.) (clinician's multiple bases for opinion inadequate to support causation opinion), *aff'd,* 46 F.3d 1120 (3d Cir. 1994); *Landrigan*, 605 A.2d at 1083–89 (permitting both clinicians and epidemiologists to testify to specific causation provided the methodology used is sound); Trach v. Fellin, 817 A.2d 1102, 1118–19 (Pa. Super. Ct. 2003) (toxicologist and pathologist permitted to testify to specific causation).

have grappled.[182] The remainder of this section is predominantly an explanation of judicial opinions. It is, in addition, in its discussion of the reasoning behind applying the risk estimates of an epidemiologic body of evidence to an individual, informed by epidemiologic principles and methodological research.

Before proceeding, one more caveat is in order. This section assumes that epidemiologic evidence has been used as proof of causation for a given plaintiff. The discussion does not address whether a plaintiff must use epidemiologic evidence to prove causation.[183]

Two legal issues arise with regard to the role of epidemiology in proving individual causation: admissibility and sufficiency of evidence to meet the burden of production. The first issue tends to receive less attention by the courts but nevertheless deserves mention. An epidemiologic study that is sufficiently rigorous to justify a conclusion that it is scientifically valid should be admissible,[184] as it tends to make an issue in dispute more or less likely.[185]

---

182. *See* Restatement (Third) of Torts: Liability for Physical and Emotional Harm § 28 cmt. c(3) (2010) ("Scientists who conduct group studies do not examine specific causation in their research. No scientific methodology exists for assessing specific causation for an individual based on group studies. Nevertheless, courts have reasoned from the preponderance-of-the-evidence standard to determine the sufficiency of scientific evidence on specific causation when group-based studies are involved").

183. *See id.* § 28 cmt. c(3) & rptrs. note ("most courts have appropriately declined to impose a threshold requirement that a plaintiff always must prove causation with epidemiologic evidence"); *see also* Westberry v. Gislaved Gummi AB, 178 F.2d 257 (4th Cir. 1999) (acute response, differential diagnosis ruled out other known causes of disease, dechallenge, rechallenge tests by expert that were consistent with exposure to defendant's agent causing disease, and absence of epidemiologic or toxicologic studies; holding that expert's testimony on causation was properly admitted); Zuchowicz v. United States, 140 F.3d 381 (2d Cir. 1998); *In re* Heparin Prods. Liab. Litig. 2011 WL 2971918, at *7-10 (N.D. Ohio July 21, 2011).

184. *See* DeLuca v. Merrell Dow Pharms., Inc., 911 F.2d 941, 958 (3d Cir. 1990); *cf.* Kehm v. Procter & Gamble Co., 580 F. Supp. 890, 902 (N.D. Iowa 1982) ("These [epidemiologic] studies were highly probative on the issue of causation—they all concluded that an association between tampon use and menstrually related TSS [toxic shock syndrome] cases exists."), *aff'd,* 724 F.2d 613 (8th Cir. 1984).

Hearsay concerns may limit the independent admissibility of the study, but the study could be relied on by an expert in forming an opinion and may be admissible pursuant to Fed. R. Evid. 703 as part of the underlying facts or data relied on by the expert.

In *Ellis v. International Playtex, Inc.,* 745 F.2d 292, 303 (4th Cir. 1984), the court concluded that certain epidemiologic studies were admissible despite criticism of the methodology used in the studies. The court held that the claims of bias went to the studies' weight rather than their admissibility. *Cf.* Christophersen v. Allied-Signal Corp., 939 F.2d 1106, 1109 (5th Cir. 1991) ("As a general rule, questions relating to the bases and sources of an expert's opinion affect the weight to be assigned that opinion rather than its admissibility. . . . ").

185. Even if evidence is relevant, it may be excluded if its probative value is substantially outweighed by prejudice, confusion, or inefficiency. Fed. R. Evid. 403. However, exclusion of an otherwise relevant epidemiologic study on Rule 403 grounds is unlikely.

In *Daubert v. Merrell Dow Pharmaceuticals, Inc.,* 509 U.S. 579, 591 (1993), the Court invoked the concept of "fit," which addresses the relationship of an expert's scientific opinion to the facts of the case and the issues in dispute. In a toxic substance case in which cause in fact is disputed, an epi-

Far more courts have confronted the role that epidemiology plays with regard to the sufficiency of the evidence and the burden of production.[186] The civil burden of proof is described most often as requiring belief by the factfinder "that what is sought to be proved is more likely true than not true."[187] The relative risk from epidemiologic studies can be adapted to this 50%-plus standard to yield a probability or likelihood that an agent caused an individual's disease.[188] An important caveat is necessary, however. The discussion below speaks in terms of the magnitude of the relative risk or association found in a study. However, before an association or relative risk is used to make a statement about the probability of individual causation, the inferential judgment, described in Section V, that the association is truly causal rather than spurious, is required: "[A]n agent cannot be considered to cause the illness of a specific person unless it is recognized as a cause of that disease in general."[189] The following discussion should be read with this caveat in mind.[190]

demiologic study of the same agent to which the plaintiff was exposed that examined the association with the same disease from which the plaintiff suffers would undoubtedly have sufficient "fit" to be a part of the basis of an expert's opinion. The Court's concept of "fit," borrowed from *United States v. Downing,* 753 F.2d 1224, 1242 (3d Cir. 1985), appears equivalent to the more familiar evidentiary concept of probative value, albeit one requiring assessment of the scientific reasoning the expert used in drawing inferences from methodology or data to opinion.

186. We reiterate a point made at the outset of this section: This discussion of the use of a threshold relative risk for specific causation is not epidemiology or an inquiry an epidemiologist would undertake. This is an effort by courts and commentators to adapt the legal standard of proof to the available scientific evidence. See *supra* text accompanying notes 175–179. While strength of association is a guideline for drawing an inference of causation from an association, *see supra* Section V, there is no specified threshold required.

187. Kevin F. O'Malley et al., Federal Jury Practice and Instructions § 104.01 (5th ed. 2000); *see also* United States v. Fatico, 458 F. Supp. 388, 403 (E.D.N.Y. 1978) ("Quantified, the preponderance standard would be 50%+ probable."), *aff'd,* 603 F.2d 1053 (2d Cir. 1979).

188. An adherent of the frequentist school of statistics would resist this adaptation, which may explain why many epidemiologists and toxicologists also resist it. To take the step identified in the text of using an epidemiologic study outcome to determine the probability of specific causation requires a shift from a frequentist approach, which involves sampling or frequency data from an empirical test, to a subjective probability about a discrete event. Thus, a frequentist might assert, after conducting a sampling test, that 60% of the balls in an opaque container are blue. The same frequentist would resist the statement, "The probability that a single ball removed from the box and hidden behind a screen is blue is 60%." The ball is either blue or not, and no frequentist data would permit the latter statement. "[T]here is no logically rigorous definition of what a statement of probability means with reference to an individual instance. . . ." Lee Loevinger, *On Logic and Sociology,* 32 Jurimetrics J. 527, 530 (1992); *see also* Steve Gold, *Causation in Toxic Torts: Burdens of Proof, Standards of Persuasion and Statistical Evidence,* 96 Yale L.J. 376, 382–92 (1986). Subjective probabilities about unique events are employed by those using Bayesian methodology. *See* Kaye, *supra* note 80, at 54–62; David H. Kaye & David A. Freedman, Reference Guide on Statistics, Section IV.D, in this manual.

189. Cole, *supra* note 65, at 10,284.

190. We emphasize this caveat, both because it is not intuitive and because some courts have failed to appreciate the difference between an association and a causal relationship. *See, e.g.,* Forsyth v. Eli Lilly & Co., Civ. No. 95-00185 ACK, 1998 U.S. Dist. LEXIS 541, at *26–*31 (D. Haw. Jan. 5, 1998). *But see*

Some courts have reasoned that when epidemiologic studies find that exposure to the agent causes an incidence in the exposed group that is more than twice the incidence in the unexposed group (i.e., a relative risk greater than 2.0), the probability that exposure to the agent caused a similarly situated individual's disease is greater than 50%.[191] These courts, accordingly, hold that when there is group-based evidence finding that exposure to an agent causes an incidence of disease in the exposed group that is more than twice the incidence in the unexposed group, the evidence is sufficient to satisfy the plaintiff's burden of production and permit submission of specific causation to a jury. In such a case, the factfinder may find that it is more likely than not that the substance caused the particular plaintiff's disease. Courts, thus, have permitted expert witnesses to testify to specific causation based on the logic of the effect of a doubling of the risk.[192]

While this reasoning has a certain logic as far as it goes, there are a number of significant assumptions and important caveats that require explication:

1. *A valid study and risk estimate*. The propriety of this "doubling" reasoning depends on group studies identifying a genuine causal relationship and a reasonably reliable measure of the increased risk.[193] This requires attention

Berry v. CSX Transp., Inc., 709 So. 2d 552, 568 (Fla. Dist. Ct. App. 1998) ("From epidemiologic studies demonstrating an association, an epidemiologist may or may not infer that a causal relationship exists.").

191. An alternative, yet similar, means to address probabilities in individual cases is use of the attributable fraction parameter, also known as the attributable risk. *See supra* Section III.C. The attributable fraction is that portion of the excess risk that can be attributed to an agent, above and beyond the background risk that is due to other causes. Thus, when the relative risk is greater than 2.0, the attributable fraction exceeds 50%.

192. For a comprehensive list of cases that support proof of causation based on group studies, see Restatement (Third) of Torts: Liability for Physical and Emotional Harm § 28 cmt. c(4) rptrs. note (2010). The Restatement catalogues those courts that require a relative risk in excess of 2.0 as a threshold for sufficient proof of specific causation and those courts that recognize that a lower relative risk than 2.0 can support specific causation, as explained below. Despite considerable disagreement on whether a relative risk of 2.0 is required or merely a taking-off point for determining the sufficiency of the evidence on specific causation, two commentators who surveyed the cases observed that "[t]here were no clear differences in outcomes as between federal and state courts." Russellyn S. Carruth & Bernard D. Goldstein, *Relative Risk Greater than Two in Proof of Causation in Toxic Tort Litigation*, 41 Jurimetrics J. 195, 199 (2001).

193. Indeed, one commentator contends that, because epidemiology is sufficiently imprecise to accurately measure small increases in risk, in general, studies that find a relative risk less than 2.0 should not be sufficient for causation. The concern is not with specific causation but with general causation and the likelihood that an association less than 2.0 is noise rather than reflecting a true causal relationship. *See* Michael D. Green, *The Future of Proportional Liability, in* Exploring Tort Law (Stuart Madden ed., 2005); *see also* Samuel M. Lesko & Allen A. Mitchell, *The Use of Randomized Controlled Trials for Pharmacoepidemiology Studies, in* Pharmacoepidemiology 599, 601 (Brian L. Strom ed., 4th ed. 2005) ("it is advisable to use extreme caution in making causal inferences from small relative risks derived from observational studies"); Gary Taubes, *Epidemiology Faces Its Limits*, 269 Science 164 (1995) (explaining views of several epidemiologists about a threshold relative risk of 3.0 to seriously consider a causal relationship); N.E. Breslow & N.E. Day, *Statistical Methods in Cancer Research, in* The Analysis

to the possibility of random error, bias, or confounding being the source of the association rather than a true causal relationship as explained in Sections IV and V, *supra*.[194]

2. *Similarity among study subjects and plaintiff.* Only if the study subjects and the plaintiff are similar with respect to other risk factors will a risk estimate from a study or studies be valid when applied to an individual.[195] Thus, if those exposed in a study of the risk of lung cancer from smoking smoked half a pack of cigarettes a day for 20 years, the degree of increased incidence of lung cancer among them cannot be extrapolated to someone who smoked two packs of cigarettes for 30 years without strong (and questionable) assumptions about the dose–response relationship.[196] This is also applicable to risk factors for competing causes. Thus, if all of the subjects in a study are participating because they were identified as having a family history of heart disease, the magnitude of risk found in a study of smok-

---

of Case-Control Studies 36 (IARC Pub. No. 32, 1980) ("[r]elative risks of less than 2.0 may readily reflect some unperceived bias or confounding factor"); David A. Freedman & Philip B. Stark, *The Swine Flu Vaccine and Guillain-Barré Syndrome: A Case Study in Relative Risk and Specific Causation,* 64 Law & Contemp. Probs. 49, 61 (2001) ("If the relative risk is near 2.0, problems of bias and confounding in the underlying epidemiologic studies may be serious, perhaps intractable.").

194. An excellent explanation for why differential diagnoses generally are inadequate without further proof of general causation was provided in *Cavallo v. Star Enterprises,* 892 F. Supp. 756 (E.D. Va. 1995), *aff'd in relevant part,* 100 F.3d 1150 (4th Cir. 1996):

> The process of differential diagnosis is undoubtedly important to the question of "specific causation". If other possible causes of an injury cannot be ruled out, or at least the probability of their contribution to causation minimized, then the "more likely than not" threshold for proving causation may not be met. But, it is also important to recognize that a fundamental assumption underlying this method is that the final, suspected "cause" remaining after this process of elimination must actually be capable of causing the injury. That is, the expert must "rule in" the suspected cause as well as "rule out" other possible causes. And, of course, expert opinion on this issue of "general causation" must be derived from a scientifically valid methodology.

Id. at 771 (footnote omitted); *see also* Ruggiero v. Warner-Lambert Co., 424 F.3d 249, 254 (2d Cir. 2005); Norris v. Baxter Healthcare Corp., 397 F.3d 878, 885 (10th Cir. 2005); Meister v. Med. Eng'g Corp., 267 F.3d 1123, 1128–29 (D.C. Cir. 2001); Bickel v. Pfizer, Inc., 431 F. Supp. 2d 918, 923–24 (N.D. Ind. 2006); *In re* Rezulin Prods. Liab. Litig., 369 F. Supp. 2d 398, 436 (S.D.N.Y. 2005); Coastal Tankships, U.S.A., Inc. v. Anderson, 87 S.W.3d 591, 608–09 (Tex. Ct. App. 2002); *see generally* Joseph Sanders & Julie Machal-Fulks, *The Admissibility of Differential Diagnosis Testimony to Prove Causation in Toxic Tort Cases: The Interplay of Adjective and Substantive Law,* 64 Law & Contemp. Probs. 107, 122–25 (2001) (discussing cases rejecting differential diagnoses in the absence of other proof of general causation and contrary cases).

195 "The basic premise of probability of causation is that individual risk can be determined from epidemiologic data for a representative population; however the premise only holds if the individual is truly representative of the reference population." Council on Scientific Affairs, American Medical Association, *Radioepidemiological Tables* 257 JAMA 806 (1987).

196. Conversely, a risk estimate from a study that involved a greater exposure is not applicable to an individual exposed to a lower dose. *See, e.g., In re* Bextra & Celebrex Mktg. Sales Practices & Prod. Liab. Litig., 524 F. Supp. 2d 1166, 1175–76 (N.D. Cal. 2007) (relative risk found in studies of those who took twice the dose of others could not support expert's opinion of causation for latter group).

ing on the risk of heart disease cannot validly be applied to an individual without such a family history. Finally, if an individual has been differentially exposed to other risk factors from those in a study, the results of the study will not provide an accurate basis for the probability of causation for the individual.[197] Consider once again a study of the effect of smoking on lung cancer among subjects who have no asbestos exposure. The relative risk of smoking in that study would not be applicable to an asbestos insulation worker. More generally, if the study subjects are heterogeneous with regard to risk factors related to the outcome of interest, the relative risk found in a study represents *an average risk for the group* rather than a uniform increased risk applicable to each individual.[198]

3. *Nonacceleration of disease.* Another assumption embedded in using the risk findings of a group study to determine the probability of causation in an individual is that the disease is one that never would have been contracted absent exposure. Put another way, the assumption is that the agent did not merely accelerate occurrence of the disease without affecting the lifetime risk of contracting the disease. Birth defects are an example of an outcome that is not accelerated. However, for most of the chronic diseases of adulthood, it is not possible for epidemiologic studies to distinguish between acceleration of disease and causation of new disease. If, in fact, acceleration

197. *See* David H. Kaye & David A. Freedman, Reference Guide on Statistics, in this manual (explaining the problems of employing a study outcome to determine the probability of an individual's having contracted the disease from exposure to the agent because of variations in individuals that bear on the risk of a given individual contracting the disease); David A. Freedman & Philip Stark, *The Swine Flu Vaccine and Guillain-Barré Syndrome: A Case Study in Relative Risk and Specific Causation,* 23 Evaluation Rev. 619 (1999) (analyzing the role that individual variation plays in determining the probability of specific causation based on the relative risk found in a study and providing a mathematical model for calculating the effect of individual variation); Mark Parascandola, *What Is Wrong with the Probability of Causation?* 39 Jurimetrics J. 29 (1998).

198. The comment of two prominent epidemiologists on this subject is illuminating:

> We cannot measure the individual risk, and assigning the average value to everyone in the category reflects nothing more than our ignorance about the determinants of lung cancer that interact with cigarette smoke. It is apparent from epidemiological data that some people can engage in chain smoking for many decades without developing lung cancer. Others are or will become primed by unknown circumstances and need only to add cigarette smoke to the nearly sufficient constellation of causes to initiate lung cancer. In our ignorance of these hidden causal components, the best we can do in assessing risk is to classify people according to measured causal risk indicators and then assign the average observed within a class to persons within the class.

Rothman & Greenland, *supra* note 131, at 9; *see also* Ofer Shpilberg et al., *The Next Stage: Molecular Epidemiology,* 50 J. Clinical Epidemiology 633, 637 (1997) ("A 1.5-fold relative risk may be composed of a 5-fold risk in 10% of the population, and a 1.1-fold risk in the remaining 90%, or a 2-fold risk in 25% and a 1.1-fold for 75%, or a 1.5-fold risk for the entire population.").

is involved, the relative risk from a study will understate the probability that exposure accelerated the occurrence of the disease.[199]

4. *Agent operates independently*. Employing a risk estimate to determine the probability of causation is not valid if the agent interacts with another cause in a way that results in an increase in disease beyond merely the sum of the increased incidence due to each agent separately. For example, the relative risk of lung cancer due to smoking is around 10, while the relative risk for asbestos exposure is approximately 5. The relative risk for someone exposed to both is not the arithmetic sum of the two relative risks, that is, 15, but closer to the product (50- to 60-fold), reflecting an interaction between the two.[200] Neither of the individual agent's relative risks can be employed to estimate the probability of causation in someone exposed to both asbestos and cigarette smoke.[201]

5. *Other assumptions*. Additional assumptions include (a) the agent of interest is not responsible for fatal diseases other than the disease of interest[202] and (b) the agent does not provide a protective effect against the outcome of interest in a subpopulation of those being studied.[203]

Evidence in a given case may challenge one or more of these assumptions. Bias in a study may suggest that the study findings are inaccurate and should be estimated to be higher or lower or, even, that the findings are spurious, that is, do not reflect a true causal relationship. A plaintiff may have been exposed to a

---

199. *See* Sander Greenland & James M. Robins, *Epidemiology, Justice, and the Probability of Causation,* 40 Jurimetrics J. 321 (2000); 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). If acceleration occurs, then the appropriate characterization of the harm for purposes of determining damages would have to be addressed. A defendant who only accelerates the occurrence of harm, say, chronic back pain, that would have occurred independently in the plaintiff at a later time is not liable for the same amount of damages as a defendant who causes a lifetime of chronic back pain. *See* David A. Fischer, *Successive Causes and the Enigma of Duplicated Harm,* 66 Tenn. L. Rev. 1127, 1127 (1999); Michael D. Green, *The Intersection of Factual Causation and Damages,* 55 DePaul L. Rev. 671 (2006).

200. We use interaction to mean that the combined effect is other than the additive sum of each effect, which is what we would expect if the two agents operate independently. Statisticians employ the term interaction in a different manner to mean the outcome deviates from what was expected in the model specified in advance. *See* Jay S. Kaufman, *Interaction Reaction,* 20 Epidemiology 159 (2009); Sander Greenland & Kenneth J. Rothman, *Concepts of Interaction, in* Rothman & Greenland, *supra* note 131, at 329.

201. *See* Restatement (Third) of Torts: Liability for Physical and Emotional Harm § 28 cmt. c(5) (2010); Jan Beyea & Sander Greenland, *The Importance of Specifying the Underlying Biologic Model in Estimating the Probability of Causation,* 76 Health Physics 269 (1999).

202. This is because in the epidemiologic studies relied on, those deaths caused by the alternative disease process will mask the true magnitude of increased incidence of the studied disease when the study subjects die before developing the disease of interest.

203. *See* Greenland & Robins, *supra,* note 198, at 332–33.

dose of the agent in question that is greater or lower than that to which those in the study were exposed.[204] A plaintiff may have individual factors, such as higher age than those in the study, that make it less likely that exposure to the agent caused the plaintiff's disease. Similarly, an individual plaintiff may be able to rule out other known (background) causes of the disease, such as genetics, that increase the likelihood that the agent was responsible for that plaintiff's disease. Evidence of a pathological mechanism may be available for the plaintiff that is relevant to the cause of the plaintiff's disease.[205] Before any causal relative risk from an epidemiologic study can be used to estimate the probability that the agent in question caused an individual plaintiff's disease, consideration of these (and related) factors is required.[206]

Having additional evidence that bears on individual causation has led a few courts to conclude that a plaintiff may satisfy his or her burden of production even if a relative risk less than 2.0 emerges from the epidemiologic evidence.[207] For example, genetics might be known to be responsible for 50% of the incidence of a disease independent of exposure to the agent.[208] If genetics can be ruled out

---

204. *See supra* Section V.C; *see also* Ferebee v. Chevron Chem. Co., 736 F.2d 1529, 1536 (D.C. Cir. 1984) ("The dose–response relationship at low levels of exposure for admittedly toxic chemicals like paraquat is one of the most sharply contested questions currently being debated in the medical community."); *In re* Joint E. & S. Dist. Asbestos Litig., 774 F. Supp. 113, 115 (S.D.N.Y. 1991) (discussing different relative risks associated with different doses), *rev'd on other grounds,* 964 F.2d 92 (2d Cir. 1992).

205. *See* Tobin v. Astra Pharm. Prods., Inc., 993 F.2d 528 (6th Cir. 1993) (plaintiff's expert relied predominantly on pathogenic evidence).

206. *See* Merrell Dow Pharms., Inc. v. Havner, 953 S.W.2d 706, 720 (Tex. 1997); Smith v. Wyeth-Ayerst Labs. Co., 278 F. Supp. 2d 684, 708–09 (W.D.N.C. 2003) (describing expert's effort to refine relative risk applicable to plaintiff based on specific risk characteristics applicable to her, albeit in an ill-explained manner); McDarby v. Merck & Co., 949 A.2d 223 (N.J. Super. Ct. App. Div. 2008); Mary Carter Andrues, *Proof of Cancer Causation in Toxic Waste Litigation,* 61 S. Cal. L. Rev. 2075, 2100–04 (1988). An example of a judge sitting as factfinder and considering individual factors for a number of plaintiffs in deciding cause in fact is contained in *Allen v. United States,* 588 F. Supp. 247, 429–43 (D. Utah 1984), *rev'd on other grounds,* 816 F.2d 1417 (10th Cir. 1987), *cert. denied,* 484 U.S. 1004 (1988); *see also* Manko v. United States, 636 F. Supp. 1419, 1437 (W.D. Mo. 1986), *aff'd,* 830 F.2d 831 (8th Cir. 1987).

207. *In re* Hanford Nuclear Reservation Litig., 292 F.3d 1124, 1137 (9th Cir. 2002) (applying Washington law) (recognizing the role of individual factors that may modify the probability of causation based on the relative risk); Magistrini v. One Hour Martinizing Dry Cleaning, 180 F. Supp. 2d 584, 606 (D.N.J. 2002) ("[A] relative risk of 2.0 is not so much a password to a finding of causation as one piece of evidence, among others for the court to consider in determining whether an expert has employed a sound methodology in reaching his or her conclusion."); Miller v. Pfizer, Inc., 196 F. Supp. 2d 1062, 1079 (D. Kan. 2002) (rejecting a threshold of 2.0 for the relative risk and recognizing that even a relative risk greater than 2.0 may be insufficient); Pafford v. Sec'y, Dept. of Health & Human Servs., 64 Fed. Cl. 19 (2005) (acknowledging that epidemiologic studies finding a relative risk of less than 2.0 can provide supporting evidence of causation), *aff'd,* 451 F.3d 1352 (Fed. Cir. 2006).

208. *See generally* Steve C. Gold, *The More We Know, the Less Intelligent We Are? How Genomic Information Should, and Should Not, Change Toxic Tort Causation Doctrine,* 34 Harv. Envtl. L. Rev. 369 (2010); Jamie A. Grodsky, *Genomics and Toxic Torts: Dismantling the Risk-Injury Divide,* 59 Stan. L. Rev.

in an individual's case, then a relative risk greater than 1.5 might be sufficient to support an inference that the agent was more likely than not responsible for the plaintiff's disease.[209]

Indeed, this idea of eliminating a known and competing cause is central to the methodology popularly known in legal terminology as differential diagnosis[210] but is more accurately referred to as differential etiology.[211] Nevertheless, the logic is sound if the label is not: Eliminating other known and competing causes increases the probability that a given individual's disease was caused by exposure to the agent. In a differential etiology, an expert first determines other known causes of the disease in question and then attempts to ascertain whether those competing causes can be "ruled out" as a cause of plaintiff's disease[212] as in the

1671 (2007); Gary E. Marchant, *Genetic Data in Toxic Tort Litigation,* 14 J.L. & Pol'y 7 (2006); Gary E. Marchant, *Genetics and Toxic Torts,* 31 Seton Hall L. Rev. 949 (2001).

209. The use of probabilities in excess of .50 to support a verdict results in an all-or-nothing approach to damages that some commentators have criticized. The criticism reflects the fact that defendants responsible for toxic agents with a relative risk just above 2.0 may be required to pay damages not only for the disease that their agents caused, but also for all instances of the disease. Similarly, those defendants whose agents increase the risk of disease by less than a doubling may not be required to pay damages for any of the disease that their agents caused. *See, e.g.,* 2 American Law Inst., Reporter's Study on Enterprise Responsibility for Personal Injury: Approaches to Legal and Institutional Change 369–75 (1991). Judge Posner has been in the vanguard of those advocating that damages be awarded on a proportional basis that reflects the probability of causation or liability. *See, e.g.,* Doll v. Brown, 75 F.3d 1200, 1206–07 (7th Cir. 1996). To date, courts have not adopted a rule that would apportion damages based on the probability of cause in fact in toxic substances cases. *See* Green, *supra* note 192.

210. Physicians regularly employ differential diagnoses in treating their patients to identify the disease from which the patient is suffering. *See* Jennifer R. Jamison, Differential Diagnosis for Primary Practice (1999).

211. It is important to emphasize that the term "differential diagnosis" in a clinical context refers to identifying a set of diseases or illnesses responsible for the patient's symptoms, while "differential etiology" refers to identifying the causal factors involved in an individual's disease or illness. For many health conditions, the *cause* of the disease or illness has no relevance to its treatment, and physicians, therefore, do not employ this term or pursue that question. *See* Zandi v. Wyeth a/k/a Wyeth, Inc., No. 27-CV-06-6744, 2007 WL 3224242 (Minn. Dist. Ct. Oct. 15, 2007) (commenting that physicians do not attempt to determine the cause of breast cancer). Thus, the standard differential diagnosis performed by a physician is not to determine the cause of a patient's disease. *See* John B. Wong et al., Reference Guide on Medical Testimony, in this manual; Edward J. Imwinkelried, *The Admissibility and Legal Sufficiency of Testimony About Differential Diagnosis (Etiology): of Under — and Over — Estimations,* 56 Baylor L. Rev. 391, 402–03 (2004); *see also* Turner v. Iowa Fire Equip. Co., 229 F.3d 1202, 1208 (8th Cir. 2000) (distinguishing between differential diagnosis conducted for the purpose of identifying the disease from which the patient suffers and one attempting to determine the cause of the disease); Creanga v. Jardal, 886 A.2d 633, 639 (N.J. 2005) ("Whereas most physicians use the term to describe the process of determining which of several diseases is causing a patient's *symptoms*, courts have used the term in a more general sense to describe the process by which causes of the patient's *condition* are identified.").

212. Courts regularly affirm the legitimacy of employing differential diagnostic methodology. *See, e.g., In re* Ephedra Prods. Liab. Litig., 393 F. Supp. 2d 181, 187 (S.D.N.Y. 2005); Easum v. Miller, 92 P.3d 794, 802 (Wyo. 2004) ("Most circuits have held that a reliable differential diagnosis satisfies *Daubert* and provides a valid foundation for admitting an expert opinion. The circuits reason that a differential diagnosis is a tested methodology, has been subjected to peer review/publication, does not

genetics example in the preceding paragraph. Similarly, an expert attempting to determine whether an individual's emphysema was caused by occupational chemical exposure would inquire whether the individual was a smoker. By ruling out (or ruling in) the possibility of other causes, the probability that a given agent was the cause of an individual's disease can be refined. Differential etiologies are most critical when the agent at issue is relatively weak and is not responsible for a large proportion of the disease in question.

Although differential etiologies are a sound methodology in principle, this approach is only valid if general causation exists and a substantial proportion of competing causes are known.[213] Thus, for diseases for which the causes are largely unknown, such as most birth defects, a differential etiology is of little benefit.[214] And, like any scientific methodology, it can be performed in an unreliable manner.[215]

VIII. Acknowledgments

---

frequently lead to incorrect results, and is generally accepted in the medical community." (quoting Turner v. Iowa Fire Equip. Co., 229 F.3d 1202, 1208 (8th Cir. 2000))); Alder v. Bayer Corp., AGFA Div., 61 P.3d 1068, 1084–85 (Utah 2002).

213. Courts have long recognized that to prove causation plaintiff need not eliminate *all* potential competing causes. *See* Stubbs v. City of Rochester, 134 N.E. 137, 140 (N.Y. 1919) (rejecting defendant's argument that plaintiff was required to eliminate all potential competing causes of typhoid); *see also* Easum v. Miller, 92 P.3d 794, 804 (Wyo. 2004). At the same time, before a competing cause should be considered relevant to a differential diagnosis, there must be adequate evidence that it *is* a cause of the disease. *See* Cooper v. Smith & Nephew, Inc., 259 F.3d 194, 202 (4th Cir. 2001); Ranes v. Adams Labs., Inc., 778 N.W.2d 677, 690 (Iowa 2010).

214. *See* Perry v. Novartis Pharms. Corp., 564 F. Supp. 2d 452, 469 (E.D. Pa. 2008) (finding experts' testimony inadmissible because of failure to account for idiopathic (unknown) causes in conducting differential diagnosis); Soldo v. Sandoz Pharms. Corp., 244 F. Supp. 2d 434, 480, 519 (W.D. Pa. 2003) (criticizing expert for failing to account for idiopathic causes); Magistrini v. One Hour Martinizing Dry Cleaning, 180 F. Supp. 2d 584, 609 (D.N.J. 2002) (observing that 90–95% of leukemias are of unknown causes, but proceeding incorrectly to assert that plaintiff was obliged to prove that her exposure to defendant's benzene was *the* cause of her leukemia rather than simply a cause of the disease that combined with other exposures to benzene). *But see* Ruff v. Ensign-Bickford Indus., Inc., 168 F. Supp. 2d 1271, 1286 (D. Utah 2001) (responding to defendant's evidence that most instances of disease are of unknown origin by stating that such matter went to the weight to be attributed to plaintiff's expert's testimony not its admissibility).

215. Numerous courts have concluded that, based on the manner in which a differential diagnosis was conducted, it was unreliable and the expert's testimony based on it is inadmissible. *See, e.g.,* Glastetter v. Novartis Pharms. Corp., 252 F.3d 986, 989 (8th Cir. 2001).

Glossary of Terms

The following terms and definitions were adapted from a variety of sources, including A Dictionary of Epidemiology (Miquel M. Porta et al. eds., 5th ed. 2008); 1 Joseph L. Gastwirth, Statistical Reasoning in Law and Public Policy (1988); James K. Brewer, Everything You Always Wanted to Know about Statistics, but Didn't Know How to Ask (1978); and R.A. Fisher, Statistical Methods for Research Workers (1973).

adjustment. Methods of modifying an observed association to take into account the effect of risk factors that are not the focus of the study and that distort the observed association between the exposure being studied and the disease outcome. See also direct age adjustment, indirect age adjustment.

agent. Also, risk factor. A factor, such as a drug, microorganism, chemical substance, or form of radiation, whose presence or absence can result in the occurrence of a disease. A disease may be caused by a single agent or a number of independent alternative agents, or the combined presence of a complex of two or more factors may be necessary for the development of the disease.

alpha. The level of statistical significance chosen by a researcher to determine if any association found in a study is sufficiently unlikely to have occurred by chance (as a result of random sampling error) if the null hypothesis (no association) is true. Researchers commonly adopt an alpha of .05, but the choice is arbitrary, and other values can be justified.

alpha error. Also called Type I error and false–positive error, alpha error occurs when a researcher rejects a null hypothesis when it is actually true (i.e., when there is no association). This can occur when an apparent difference is observed between the control group and the exposed group, but the difference is not real (i.e., it occurred by chance). A common error made by lawyers, judges, and academics is to equate the level of alpha with the legal burden of proof.

association. The degree of statistical relationship between two or more events or variables. Events are said to be associated when they occur more or less frequently together than one would expect by chance. Association does not necessarily imply a causal relationship. Events are said not to have an association when the agent (or independent variable) has no apparent effect on the incidence of a disease (the dependent variable). This corresponds to a relative risk of 1.0. A negative association means that the events occur less frequently together than one would expect by chance, thereby implying a preventive or protective role for the agent (e.g., a vaccine).

attributable fraction. Also, attributable risk. The proportion of disease in exposed individuals that can be attributed to exposure to an agent, as distinguished from the proportion of disease attributed to all other causes.

attributable proportion of risk (PAR). This term has been used to denote the fraction of risk that is attributable to exposure to a substance (e.g., *X* percent of lung cancer is attributable to cigarettes). Synonymous terms include attributable fraction, attributable risk, etiologic fraction, population attributable risk, and risk difference. See attributable risk.

background risk of disease. Also, background rate of disease. Rate of disease in a population that has no known exposures to an alleged risk factor for the disease. For example, the background risk for all birth defects is 3–5% of live births.

beta error. Also called Type II error and false-negative error. Occurs when a researcher fails to reject a null hypothesis when it is incorrect (i.e., when there is an association). This can occur when no statistically significant difference is detected between the control group and the exposed group, but a difference does exist.

bias. Any effect at any stage of investigation or inference tending to produce results that depart systematically from the true values. In epidemiology, the term bias does not necessarily carry an imputation of prejudice or other subjective factor, such as the experimenter's desire for a particular outcome. This differs from conventional usage, in which bias refers to a partisan point of view.

biological marker. A physiological change in tissue or body fluids that occurs as a result of an exposure to an agent and that can be detected in the laboratory. Biological markers are only available for a small number of chemicals.

biological plausibility. Consideration of existing knowledge about human biology and disease pathology to provide a judgment about the plausibility that an agent causes a disease.

case-comparison study. See case-control study.

case–control study. Also, case-comparison study, case history study, case referent study, retrospective study. A study that starts with the identification of persons with a disease (or other outcome variable) and a suitable control (comparison, reference) group of persons without the disease. Such a study is often referred to as retrospective because it starts after the onset of disease and looks back to the postulated causal factors.

case group. A group of individuals who have been exposed to the disease, intervention, procedure, or other variable whose influence is being studied.

causation. As used here, an event, condition, characteristic, or agent being a necessary element of a set of other events that can produce an outcome, such as a disease. Other sets of events may also cause the disease. For example, smoking is a necessary element of a set of events that result in lung cancer, yet there are other sets of events (without smoking) that cause lung cancer. Thus, a cause may be thought of as a necessary link in at least one causal chain that

results in an outcome of interest. Epidemiologists generally speak of causation in a group context; hence, they will inquire whether an increased incidence of a disease in a cohort was "caused" by exposure to an agent.

clinical trial. An experimental study that is performed to assess the efficacy and safety of a drug or other beneficial treatment. Unlike observational studies, clinical trials can be conducted as experiments and use randomization, because the agent being studied is thought to be beneficial.

cohort. Any designated group of persons followed or traced over a period of time to examine health or mortality experience.

cohort study. The method of epidemiologic study in which groups of individuals can be identified who are, have been, or in the future may be differentially exposed to an agent or agents hypothesized to influence the incidence of occurrence of a disease or other outcome. The groups are observed to find out if the exposed group is more likely to develop disease. The alternative terms for a cohort study (concurrent study, followup study, incidence study, longitudinal study, prospective study) describe an essential feature of the method, which is observation of the population for a sufficient number of person–years to generate reliable incidence or mortality rates in the population subsets. This generally implies study of a large population, study for a prolonged period (years), or both.

confidence interval. A range of values calculated from the results of a study within which the true value is likely to fall; the width of the interval reflects random error. Thus, if a confidence level of .95 is selected for a study, 95% of similar studies would result in the true relative risk falling within the confidence interval. The width of the confidence interval provides an indication of the precision of the point estimate or relative risk found in the study; the narrower the confidence interval, the greater the confidence in the relative risk estimate found in the study. Where the confidence interval contains a relative risk of 1.0, the results of the study are not statistically significant.

confounding factor. Also, confounder. A factor that is both a risk factor for the disease and a factor associated with the exposure of interest. Confounding refers to a situation in which an association between an exposure and outcome is all or partly the result of a factor that affects the outcome but is unaffected by the exposure.

control group. A comparison group comprising individuals who have not been exposed to the disease, intervention, procedure, or other variable whose influence is being studied.

cross–sectional study. A study that examines the relationship between disease and variables of interest as they exist in a population at a given time. A cross-sectional study measures the presence or absence of disease and other variables in each member of the study population. The data are analyzed to

determine if there is a relationship between the existence of the variables and disease. Because cross-sectional studies examine only a particular moment in time, they reflect the prevalence (existence) rather than the incidence (rate) of disease and can offer only a limited view of the causal association between the variables and disease. Because exposures to toxic agents often change over time, cross-sectional studies are rarely used to assess the toxicity of exogenous agents.

data dredging. Jargon that refers to results identified by researchers who, after completing a study, pore through their data seeking to find any associations that may exist. In general, good research practice is to identify the hypotheses to be investigated in advance of the study; hence, data dredging is generally frowned on. In some cases, however, researchers conduct exploratory studies designed to generate hypotheses for further study.

demographic study. See ecological study.

dependent variable. The outcome that is being assessed in a study based on the effect of another characteristic—the independent variable. Epidemiologic studies attempt to determine whether there is an association between the independent variable (exposure) and the dependent variable (incidence of disease).

differential misclassification. A form of bias that is due to the misclassification of individuals or a variable of interest when the misclassification varies among study groups. This type of bias occurs when, for example, it is incorrectly determined that individuals in a study are unexposed to the agent being studied when in fact they are exposed. See nondifferential misclassification.

direct adjustment. A technique used to eliminate any difference between two study populations based on age, sex, or some other parameter that might result in confounding. Direct adjustment entails comparison of the study group with a large reference population to determine the expected rates based on the characteristic, such as age, for which adjustment is being performed.

dose. Generally refers to the intensity or magnitude of exposure to an agent multiplied by the duration of exposure. Dose may be used to refer only to the intensity of exposure.

dose–response relationship. A relationship in which a change in amount, intensity, or duration of exposure to an agent is associated with a change—either an increase or a decrease—in risk of disease.

double blinding. A method used in experimental studies in which neither the individuals being studied nor the researchers know during the study whether any individual has been assigned to the exposed or control group. Double blinding is designed to prevent knowledge of the group to which the individual was assigned from biasing the outcome of the study.

ecological fallacy. Also, aggregation bias, ecological bias. An error that occurs from inferring that a relationship that exists for groups is also true for individuals. For example, if a country with a higher proportion of fishermen also has a higher rate of suicides, then inferring that fishermen must be more likely to commit suicide is an ecological fallacy.

ecological study. Also, demographic study. A study of the occurrence of disease based on data from populations, rather than from individuals. An ecological study searches for associations between the incidence of disease and suspected disease-causing agents in the studied populations. Researchers often conduct ecological studies by examining easily available health statistics, making these studies relatively inexpensive in comparison with studies that measure disease and exposure to agents on an individual basis.

epidemiology. The study of the distribution and determinants of disease or other health-related states and events in populations and the application of this study to control of health problems.

error. Random error (sampling error) is the error that is due to chance when the result obtained for a sample differs from the result that would be obtained if the entire population (universe) were studied.

etiologic factor. An agent that plays a role in causing a disease.

etiology. The cause of disease or other outcome of interest.

experimental study. A study in which the researcher directly controls the conditions. Experimental epidemiology studies (also clinical studies) entail random assignment of participants to the exposed and control groups (or some other method of assignment designed to minimize differences between the groups).

exposed, exposure. In epidemiology, the exposed group (or the exposed) is used to describe a group whose members have been exposed to an agent that may be a cause of a disease or health effect of interest, or possess a characteristic that is a determinant of a health outcome.

false-negative error. See beta error.

false-positive error. See alpha error.

followup study. See cohort study.

general causation. Issue of whether an agent increases the incidence of disease in a group and not whether the agent caused any given individual's disease. Because of individual variation, a toxic agent generally will not cause disease in every exposed individual.

generalizable. When the results of a study are applicable to populations other than the study population, such as the general population.

in vitro. Within an artificial environment, such as a test tube (e.g., the cultivation of tissue in vitro).

in vivo. Within a living organism (e.g., the cultivation of tissue in vivo).

incidence rate. The number of people in a specified population falling ill from a particular disease during a given period. More generally, the number of new events (e.g., new cases of a disease in a defined population) within a specified period of time.

incidence study. See cohort study.

independent variable. A characteristic that is measured in a study and that is suspected to have an effect on the outcome of interest (the dependent variable). Thus, exposure to an agent is measured in a cohort study to determine whether that independent variable has an effect on the incidence of disease, which is the dependent variable.

indirect adjustment. A technique employed to minimize error that might result when comparing two populations because of differences in age, sex, or another parameter that may independently affect the rate of disease in the populations. The incidence of disease in a large reference population, such as all residents of a country, is calculated for each subpopulation (based on the relevant parameter, such as age). Those incidence rates are then applied to the study population with its distribution of persons to determine the overall incidence rate for the study population, which provides a standardized mortality or morbidity ratio (often referred to as SMR).

inference. The intellectual process of making generalizations from observations. In statistics, the development of generalizations from sample data, usually with calculated degrees of uncertainty.

information bias. Also, observational bias. Systematic error in measuring data that results in differential accuracy of information (such as exposure status) for comparison groups.

interaction. When the magnitude or direction (positive or negative) of the effect of one risk factor differs depending on the presence or level of the other. In interaction, the effect of two risk factors together is different (greater or less) than the sum of their individual effects.

meta-analysis. A technique used to combine the results of several studies to enhance the precision of the estimate of the effect size and reduce the plausibility that the association found is due to random sampling error. Meta-analysis is best suited to pooling results from randomly controlled experimental studies, but if carefully performed, it also may be useful for observational studies.

misclassification bias. The erroneous classification of an individual in a study as exposed to the agent when the individual was not, or incorrectly classifying a study individual with regard to disease. Misclassification bias may exist in all study groups (nondifferential misclassification) or may vary among groups (differential misclassification).

morbidity rate. State of illness or disease. Morbidity rate may refer to either the incidence rate or prevalence rate of disease.

mortality rate. Proportion of a population that dies of a disease or of all causes. The numerator is the number of individuals dying; the denominator is the total population in which the deaths occurred. The unit of time is usually a calendar year.

model. A representation or simulation of an actual situation. This may be either (1) a mathematical representation of characteristics of a situation that can be manipulated to examine consequences of various actions; (2) a representation of a country's situation through an "average region" with characteristics resembling those of the whole country; or (3) the use of animals as a substitute for humans in an experimental system to ascertain an outcome of interest.

multivariate analysis. A set of techniques used when the variation in several variables has to be studied simultaneously. In statistics, any analytical method that allows the simultaneous study of two or more independent factors or variables.

nondifferential misclassification. Error due to misclassification of individuals or a variable of interest into the wrong category when the misclassification varies among study groups. The error may result from limitations in data collection, may result in bias, and will often produce an underestimate of the true association. See differential misclassification.

null hypothesis. A hypothesis that states that there is no true association between a variable and an outcome. At the outset of any observational or experimental study, the researcher must state a proposition that will be tested in the study. In epidemiology, this proposition typically addresses the existence of an association between an agent and a disease. Most often, the null hypothesis is a statement that exposure to Agent A does not increase the occurrence of Disease D. The results of the study may justify a conclusion that the null hypothesis (no association) has been disproved (e.g., a study that finds a strong association between smoking and lung cancer). A study may fail to disprove the null hypothesis, but that alone does not justify a conclusion that the null hypothesis has been proved.

observational study. An epidemiologic study in situations in which nature is allowed to take its course, without intervention from the investigator. For example, in an observational study the subjects of the study are permitted to determine their level of exposure to an agent.

odds ratio (OR). Also, cross-product ratio, relative odds. The ratio of the odds that a case (one with the disease) was exposed to the odds that a control (one without the disease) was exposed. For most purposes the odds ratio from a case-control study is quite similar to a risk ratio from a cohort study.

***p* (probability), *p*-value.** The *p*-value is the probability of getting a value of the test outcome equal to or more extreme than the result observed, given that the null hypothesis is true. The letter *p*, followed by the abbreviation "n.s." (not significant) means that $p > .05$ and that the association was not statistically significant at the .05 level of significance. The statement "$p < .05$" means that *p* is less than 5%, and, by convention, the result is deemed statistically significant. Other significance levels can be adopted, such as .01 or .1. The lower the *p*-value, the less likely that random error would have produced the observed relative risk if the true relative risk is 1.

pathognomonic. When an agent must be present for a disease to occur. Thus, asbestos is a pathognomonic agent for asbestosis. See signature disease.

placebo controlled. In an experimental study, providing an inert substance to the control group, so as to keep the control and exposed groups ignorant of their status.

power. The probability that a difference of a specified amount will be detected by the statistical hypothesis test, given that a difference exists. In less formal terms, power is like the strength of a magnifying lens in its capability to identify an association that truly exists. Power is equivalent to one minus Type II error. This is sometimes stated as Power $= 1 - \beta$.

prevalence. The percentage of persons with a disease in a population at a specific point in time.

prospective study. A study in which two groups of individuals are identified: (1) individuals who have been exposed to a risk factor and (2) individuals who have not been exposed. Both groups are followed for a specified length of time, and the proportion that develops disease in the first group is compared with the proportion that develops disease in the second group. See cohort study.

random. The term implies that an event is governed by chance. See randomization.

randomization. Assignment of individuals to groups (e.g., for experimental and control regimens) by chance. Within the limits of chance variation, randomization should make the control group and experimental group similar at the start of an investigation and ensure that personal judgment and prejudices of the investigator do not influence assignment. Randomization should not be confused with haphazard assignment. Random assignment follows a predetermined plan that usually is devised with the aid of a table of random numbers. Randomization cannot ethically be used where the exposure is known to cause harm (e.g., cigarette smoking).

randomized trial. See clinical trial.

recall bias. Systematic error resulting from differences between two groups in a study in accuracy of memory. For example, subjects who have a disease may recall exposure to an agent more frequently than subjects who do not have the disease.

relative risk (RR). The ratio of the risk of disease or death among people exposed to an agent to the risk among the unexposed. For instance, if 10% of all people exposed to a chemical develop a disease, compared with 5% of people who are not exposed, the disease occurs twice as frequently among the exposed people. The relative risk is 10%/5% = 2. A relative risk of 1 indicates no association between exposure and disease.

research design. The procedures and methods, predetermined by an investigator, to be adhered to in conducting a research project.

risk. A probability that an event will occur (e.g., that an individual will become ill or die within a stated period of time or by a certain age).

risk difference (RD). The difference between the proportion of disease in the exposed population and the proportion of disease in the unexposed population. $-1.0 \leq RD \geq 1.0$.

sample. A selected subset of a population. A sample may be random or nonrandom.

sample size. The number of subjects who participate in a study.

secular–trend study. Also, time-line study. A study that examines changes over a period of time, generally years or decades. Examples include the decline of tuberculosis mortality and the rise, followed by a decline, in coronary heart disease mortality in the United States in the past 50 years.

selection bias. Systematic error that results from individuals being selected for the different groups in an observational study who have differences other than the ones that are being examined in the study.

sensitivity. Measure of the accuracy of a diagnostic or screening test or device in identifying disease (or some other outcome) when it truly exists. For example, assume that we know that 20 women in a group of 1000 women have cervical cancer. If the entire group of 1000 women is tested for cervical cancer and the screening test only identifies 15 (of the known 20) cases of cervical cancer, the screening test has a sensitivity of 15/20, or 75%. Also see specificity.

signature disease. A disease that is associated uniquely with exposure to an agent (e.g., asbestosis and exposure to asbestos). See also pathognomonic.

significance level. A somewhat arbitrary level selected to minimize the risk that an erroneous positive study outcome that is due to random error will be accepted as a true association. The lower the significance level selected, the less likely that false–positive error will occur.

specific causation. Whether exposure to an agent was responsible for a given individual's disease.

specificity. Measure of the accuracy of a diagnostic or screening test in identifying those who are disease-free. Once again, assume that 980 women out of a group of 1000 women do not have cervical cancer. If the entire group of 1000 women is screened for cervical cancer and the screening test only iden-

tifies 900 women without cervical cancer, the screening test has a specificity of 900/980, or 92%.

standardized morbidity ratio (SMR). The ratio of the incidence of disease observed in the study population to the incidence of disease that would be expected if the study population had the same incidence of disease as some selected reference population.

standardized mortality ratio (SMR). The ratio of the incidence of death observed in the study population to the incidence of death that would be expected if the study population had the same incidence of death as some selected standard or known population.

statistical significance. A term used to describe a study result or difference that exceeds the Type I error rate (or *p*-value) that was selected by the researcher at the outset of the study. In formal significance testing, a statistically significant result is unlikely to be the result of random sampling error and justifies rejection of the null hypothesis. Some epidemiologists believe that formal significance testing is inferior to using a confidence interval to express the results of a study. Statistical significance, which addresses the role of random sampling error in producing the results found in the study, should not be confused with the importance (for public health or public policy) of a research finding.

stratification. Separating a group into subgroups based on specified criteria, such as age, gender, or socioeconomic status. Stratification is used both to control for the possibility of confounding (by separating the studied populations based on the suspected confounding factor) and when there are other known factors that affect the disease under study. Thus, the incidence of death increases with age, and a study of mortality might use stratification of the cohort and control groups based on age.

study design. See research design.

systematic error. See bias.

teratogen. An agent that produces abnormalities in the embryo or fetus by disturbing maternal health or by acting directly on the fetus in utero.

teratogenicity. The capacity for an agent to produce abnormalities in the embryo or fetus.

threshold phenomenon. A certain level of exposure to an agent below which disease does not occur and above which disease does occur.

time–line study. See secular-trend study.

toxicology. The science of the nature and effects of poisons. Toxicologists study adverse health effects of agents on biological organisms, such as live animals and cells. Studies of humans are performed by epidemiologists.

toxic substance. A substance that is poisonous.

true association. Also, real association. The association that really exists between exposure to an agent and a disease and that might be found by a perfect (but nonetheless nonexistent) study.

Type I error. Rejecting the null hypothesis when it is true. See alpha error.

Type II error. Failing to reject the null hypothesis when it is false. See beta error.

validity. The degree to which a measurement measures what it purports to measure; the accuracy of a measurement.

variable. Any attribute, condition, or other characteristic of subjects in a study that can have different numerical characteristics. In a study of the causes of heart disease, blood pressure and dietary fat intake are variables that might be measured.

References on Epidemiology

Causal Inferences (Kenneth J. Rothman ed., 1988).

William G. Cochran, Sampling Techniques (1977).

A Dictionary of Epidemiology (John M. Last et al. eds., 5th ed. 2008).

Anders Ahlbom & Steffan Norell, Introduction to Modern Epidemiology (2d ed. 1990).

Robert C. Elston & William D. Johnson, Basic Biostatistices for Geneticists and Epidemiologists (2008)

Encyclopedia of Epidemiology (Sarah E. Boslaugh ed., 2008).

Joseph L. Fleiss et al., Statistical Methods for Rates and Proportions (3d ed. 2003).

Leon Gordis, Epidemiology (4th ed. 2009).

Morton Hunt, How Science Takes Stock: The Story of Meta-Analysis (1997).

International Agency for Research on Cancer (IARC), Interpretation of Negative Epidemiologic Evidence for Carcinogenicity (N.J. Wald & R. Doll eds., 1985).

Harold A. Kahn & Christopher T. Sempos, Statistical Methods in Epidemiology (1989).

David E. Lilienfeld, *Overview of Epidemiology,* 3 Shepard's Expert & Sci. Evid. Q. 25 (1995).

David E. Lilienfeld & Paul D. Stolley, Foundations of Epidemiology (3d ed. 1994).

Marcello Pagano & Kimberlee Gauvreau, Principles of Biostatistics (2d ed. 2000).

Pharmacoepidemiology (Brian L. Strom ed., 4th ed. 2005).

Richard K. Riegelman & Robert A. Hirsch, Studying a Study and Testing a Test: How to Read the Health Science Literature (5th ed. 2005).

Bernard Rosner, Fundamentals of Biostatistics (6th ed. 2006).

Kenneth J. Rothman et al., Modern Epidemiology (3d ed. 2008).

David A. Savitz, Interpreting Epidemiologic Evidence: Strategies for Study Design and Analysis (2003).

James J. Schlesselman, Case-Control Studies: Design, Conduct, Analysis (1982).

Lisa M. Sullivan, Essentials of Biostatistics (2008).

Mervyn Susser, Epidemiology, Health and Society: Selected Papers (1987).

References on Law and Epidemiology

American Law Institute, Reporters' Study on Enterprise Responsibility for Personal Injury (1991).

Bert Black & David H. Hollander, Jr., *Unraveling Causation: Back to the Basics*, 3 U. Balt. J. Envtl. L. 1 (1993).

Bert Black & David Lilienfeld, *Epidemiologic Proof in Toxic Tort Litigation*, 52 Fordham L. Rev. 732 (1984).

Gerald Boston, *A Mass-Exposure Model of Toxic Causation: The Content of Scientific Proof and the Regulatory Experience,* 18 Colum. J. Envtl. L. 181 (1993).

Vincent M. Brannigan et al., *Risk, Statistical Inference, and the Law of Evidence: The Use of Epidemiological Data in Toxic Tort Cases,* 12 Risk Analysis 343 (1992).

Troyen Brennan, *Causal Chains and Statistical Links: The Role of Scientific Uncertainty in Hazardous-Substance Litigation,* 73 Cornell L. Rev. 469 (1988).

Troyen Brennan, *Helping Courts with Toxic Torts: Some Proposals Regarding Alternative Methods for Presenting and Assessing Scientific Evidence in Common Law Courts,* 51 U. Pitt. L. Rev. 1 (1989).

Philip Cole, *Causality in Epidemiology, Health Policy, and Law,* 27 Envtl. L. Rep. 10,279 (June 1997).

Comment, *Epidemiologic Proof of Probability: Implementing the Proportional Recovery Approach in Toxic Exposure Torts,* 89 Dick. L. Rev. 233 (1984).

George W. Conk, *Against the Odds: Proving Causation of Disease with Epidemiological Evidence,* 3 Shepard's Expert & Sci. Evid. Q. 85 (1995).

Carl F. Cranor, Toxic Torts: Science, Law, and the Possibility of Justice (2006).

Carl F. Cranor et al., *Judicial Boundary Drawing and the Need for Context-Sensitive Science in Toxic Torts After* Daubert v. Merrell Dow Pharmaceuticals, Inc., 16 Va. Envtl. L.J. 1 (1996).

Richard Delgado, *Beyond Sindell: Relaxation of Cause-in-Fact Rules for Indeterminate Plaintiffs,* 70 Cal. L. Rev. 881 (1982).

Michael Dore, *A Commentary on the Use of Epidemiological Evidence in Demonstrating Cause-in-Fact,* 7 Harv. Envtl. L. Rev. 429 (1983).

Jean Macchiaroli Eggen, *Toxic Torts, Causation, and Scientific Evidence After* Daubert, 55 U. Pitt. L. Rev. 889 (1994).

Daniel A. Farber, *Toxic Causation*, 71 Minn. L. Rev. 1219 (1987).

Heidi Li Feldman, Science and Uncertainty in Mass Exposure Litigation, 74 Tex. L. Rev. 1 (1995).

Stephen E. Fienberg et al., *Understanding and Evaluating Statistical Evidence in Litigation,* 36 Jurimetrics J. 1 (1995).

Joseph L. Gastwirth, Statistical Reasoning in Law and Public Policy (1988).

Herman J. Gibb, *Epidemiology and Cancer Risk Assessment*, *in* Fundamentals of Risk Analysis and Risk Management 23 (Vlasta Molak ed., 1997).

Steve Gold, Note, *Causation in Toxic Torts: Burdens of Proof, Standards of Persuasion and Statistical Evidence,* 96 Yale L.J. 376 (1986).

Leon Gordis, *Epidemiologic Approaches for Studying Human Disease in Relation to Hazardous Waste Disposal Sites,* 25 Hous. L. Rev. 837 (1988).

Michael D. Green, *Expert Witnesses and Sufficiency of Evidence in Toxic Substances Litigation: The Legacy of Agent Orange and Bendectin Litigation,* 86 Nw. U. L. Rev. 643 (1992).

Michael D. Green, *The Future of Proportional Liability, in* Exploring Tort Law (Stuart Madden ed., 2005).

Sander Greenland, *The Need for Critical Appraisal of Expert Witnesses in Epidemiology and Statistics,* 39 Wake Forest L. Rev. 291 (2004).

Khristine L. Hall & Ellen Silbergeld, *Reappraising Epidemiology: A Response to Mr. Dore*, 7 Harv. Envtl. L. Rev. 441 (1983).

Jay P. Kesan, *Drug Development: Who Knows Where the Time Goes?: A Critical Examination of the Post-*Daubert *Scientific Evidence Landscape,* 52 Food Drug Cosm. L.J. 225 (1997).

Jay P. Kesan, *An Autopsy of Scientific Evidence in a Post-*Daubert *World,* 84 Geo. L. Rev. 1985 (1996).

Constantine Kokkoris, Comment, DeLuca v. Merrell Dow Pharmaceuticals, Inc.: *Statistical Significance and the Novel Scientific Technique,* 58 Brook. L. Rev. 219 (1992).

James P. Leape, *Quantitative Risk Assessment in Regulation of Environmental Carcinogens,* 4 Harv. Envtl. L. Rev. 86 (1980).

David E. Lilienfeld, *Overview of Epidemiology*, 3 Shepard's Expert & Sci. Evid. Q. 23 (1995).

Junius McElveen, Jr., & Pamela Eddy, *Cancer and Toxic Substances: The Problem of Causation and the Use of Epidemiology,* 33 Clev. St. L. Rev. 29 (1984).

Modern Scientific Evidence: The Law and Science of Expert Testimony (David L. Faigman et al. eds., 2009–2010).

Note, *Development in the Law—Confronting the New Challenges of Scientific Evidence,* 108 Harv. L. Rev. 1481 (1995).

Susan R. Poulter, *Science and Toxic Torts: Is There a Rational Solution to the Problem of Causation?* 7 High Tech. L.J. 189 (1992).

Jon Todd Powell, Comment, *How to Tell the Truth with Statistics: A New Statistical Approach to Analyzing the Data in the Aftermath of* Daubert v. Merrell Dow Pharmaceuticals, 31 Hous. L. Rev. 1241 (1994).

Restatement (Third) of Torts: Liability for Physical and Emotional Harm § 28, cmt. c & rptrs. note (2010).

David Rosenberg, *The Causal Connection in Mass Exposure Cases: A Public Law Vision of the Tort System,* 97 Harv. L. Rev. 849 (1984).

Joseph Sanders, *The Bendectin Litigation: A Case Study in the Life-Cycle of Mass Torts,* 43 Hastings L.J. 301 (1992).

Joseph Sanders, *Scientific Validity, Admissibility, and Mass Torts After* Daubert, 78 Minn. L. Rev. 1387 (1994).

Joseph Sanders & Julie Machal-Fulks, *The Admissibility of Differential Diagnosis to Prove Causation in Toxic Tort Cases: The Interplay of Adjective and Substantive Law,* 64 L. & Contemp. Probs. 107 (2001).

Palma J. Strand, *The Inapplicability of Traditional Tort Analysis to Environmental Risks: The Example of Toxic Waste Pollution Victim Compensation,* 35 Stan. L. Rev. 575 (1983).

Richard W. Wright, *Causation in Tort Law*, 73 Cal. L. Rev. 1735 (1985).