Reference Guide on Epidemiology

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

Michael D. Green, B.S., 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., Dr.P.H., is Professor of Epidemiology, Johns Hopkins School of Public Health, and Professor of Pediatrics, Johns Hopkins School of Medicine, Baltimore, Maryland.

CONTENTS

I. Introduction, 335
II. What Different Kinds of Epidemiologic Studies Exist? 338
  A. Experimental and Observational Studies of Suspected Toxic Agents, 338
  B. The Types of Observational Study Design, 339
    1. Cohort studies, 340
    2. Case-control studies, 342
    3. Cross-sectional studies, 343
    4. Ecological studies, 344
  C. Epidemiologic and Toxicologic Studies, 345
III. How Should Results of an Epidemiologic Study Be Interpreted? 348
  A. Relative Risk, 348
  B. Odds Ratio, 350
  C. Attributable Risk, 351
  D. Adjustment for Study Groups That Are Not Comparable, 352
IV. What Sources of Error Might Have Produced a False Result? 354
  A. What Statistical Methods Exist to Evaluate the Possibility of Sampling Error? 355
    1. False positive error and statistical significance, 356
    2. False negative error, 362
    3. Power, 362
  B. What Biases May Have Contributed to an Erroneous Association? 363
    1. Selection bias, 363
    2. Information bias, 365
    3. Other conceptual problems, 369
  C. Could a Confounding Factor Be Responsible for the Study Result? 369
    1. What techniques can be used to prevent or limit confounding? 372
    2. What techniques can be used to identify confounding factors? 373
    3. What techniques can be used to control for confounding factors? 373

V. General Causation: Is an Exposure a Cause of the Disease? 374
A. Is There a Temporal Relationship? 376
B. How Strong Is the Association Between the Exposure and Disease? 376
C. Is There a Dose–Response Relationship? 377
D. Have the Results Been Replicated? 377
E. Is the Association Biologically Plausible (Consistent with Existing Knowledge)? 378
F. Have Alternative Explanations Been Considered? 378
G. What Is the Effect of Ceasing Exposure? 378
H. Does the Association Exhibit Specificity? 379
I. Are the Findings Consistent with Other Relevant Knowledge? 379
VI. What Methods Exist for Combining the Results of Multiple Studies? 380
VII. What Role Does Epidemiology Play in Proving Specific Causation? 381
Glossary of Terms, 387
References on Epidemiology, 398
References on Law and Epidemiology, 398

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 increasingly are presented with epidemiologic evidence as the basis of an expert's opinion on causation.[2] In the courtroom, epidemiologic research findings[3] are offered to establish or dispute whether exposure to an agent[4] caused a harmful effect or disease.[5] Epidemiologic evidence identifies

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 trying mass tort suits. Well-conducted studies are uniformly admitted. 2 Modern Scientific Evidence: The Law and Science of Expert Testimony § 28-1.1, at 302–03 (David L. Faigman et al. eds., 1997) [hereinafter Modern Scientific Evidence]. It is important to note that often the expert testifying before the court is not the scientist who conducted the study or series of studies. *See, e.g.,* DeLuca v. Merrell Dow Pharms., Inc., 911 F.2d 941, 953 (3d Cir. 1990) (pediatric pharmacologist expert's credentials sufficient pursuant to Fed. R. Evid. 702 to interpret epidemiologic studies and render an opinion based thereon); *cf.* Landrigan v. Celotex Corp., 605 A.2d 1079, 1088 (N.J. 1992) (epidemiologist permitted to testify to both general causation and specific causation); Loudermill v. Dow Chem. Co., 863 F.2d 566, 569 (8th Cir. 1988) (toxicologist permitted to testify that chemical caused decedent's death).

3. An epidemiologic study, which often is published in a medical journal or other scientific journal, is hearsay. An epidemiologic study that is performed by the government, such as one performed by the Centers for Disease Control (CDC), may be admissible based on the hearsay exception for government records contained in Fed. R. Evid. 803(8)(C). *See* Ellis v. International Playtex, Inc., 745 F.2d 292, 300–01 (4th Cir. 1984); Kehm v. Procter & Gamble Co., 580 F. Supp. 890, 899 (N.D. Iowa 1982), *aff'd sub nom.* Kehm v. Procter & Gamble Mfg. Co., 724 F.2d 613 (8th Cir. 1983). A study that is not conducted by the government might qualify for the learned treatise exception to the hearsay rule, Fed. R. Evid. 803(18), or possibly the catchall exceptions, Fed. R. Evid. 803(24) & 804(5). *See Ellis*, 745 F.2d at 305, 306 & n.18.

In any case, an epidemiologic study might be part of the basis of an expert's opinion and need not be independently admissible pursuant to Fed. R. Evid. 703. *See In re* "Agent Orange" Prod. Liab. Litig., 611 F. Supp. 1223, 1240 (E.D.N.Y. 1985), *aff'd*, 818 F.2d 187 (2d Cir. 1987), *cert. denied*, 487 U.S. 1234 (1988); *cf.* Grassis v. Johns-Manville Corp., 591 A.2d 671, 676 (N.J. Super. Ct. App. Div. 1991) (epidemiologic study offered in evidence to support expert's opinion under New Jersey evidentiary rule equivalent to Fed. R. Evid. 703).

4. We use *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 may also conduct studies of drugs and other pharmaceutical products to assess their efficacy and safety.

5. DeLuca v. Merrell Dow Pharms., Inc., 911 F.2d 941, 945–48, 953–59 (3d Cir. 1990) (litigation

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?).[6] 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 ten to twenty cigarettes a day had a lung cancer mortality rate that was about ten times higher than that for nonsmokers.[7] These studies identified an association between smoking cigarettes and death from lung cancer, which contributed to the determination that smoking causes lung cancer.

However, it should be emphasized that *an association is not equivalent to causation.*[8] An association identified in an epidemiologic study may or may not be causal.[9] Assessing whether an association is causal requires an understanding of

over morning sickness drug, Bendectin); Cook v. United States, 545 F. Supp. 306, 307–16 (N.D. Cal. 1982) (swine flu vaccine alleged to have caused plaintiff's Guillain-Barré disease); Allen v. United States, 588 F. Supp. 247, 416–25 (D. Utah 1984) (residents near atomic test site claimed exposure to radiation caused leukemia and other cancers), *rev'd on other grounds*, 816 F.2d 1417 (10th Cir. 1987), *cert. denied*, 484 U.S. 1004 (1988); *In re* "Agent Orange" Prod. Liab. Litig., 597 F. Supp. 740, 780–90 (E.D.N.Y. 1984) (Vietnam veterans exposed to Agent Orange and dioxin contaminant brought suit for various diseases and birth defects in their offspring), *aff'd*, 818 F.2d 145 (2d Cir. 1987); Christophersen v. Allied-Signal Corp., 939 F.2d 1106, 1115 (5th Cir. 1991) (cancer alleged to have resulted from exposure to nickel-cadmium fumes), *cert. denied*, 503 U.S. 912 (1992); Kehm v. Procter & Gamble Co., 580 F. Supp. 890, 898–902 (N.D. Iowa 1982) (toxic shock syndrome alleged to result from use of Rely tampons), *aff'd sub nom.* Kehm v. Procter & Gamble Mfg. Co., 724 F.2d 613 (8th Cir. 1983).

6. This terminology and the distinction between general causation and specific causation is widely recognized in court opinions. *See, e.g.*, Kelley v. American Heyer-Schulte Corp., 957 F. Supp. 873, 875–76 (W.D. Tex. 1997) (recognizing the different concepts of general causation and specific causation), *appeal dismissed*, 139 F.3d 899 (5th Cir. 1998); Cavallo v. Star Enter., 892 F. Supp. 756, 771 n.34 (E.D. Va. 1995), *aff'd in part and rev'd in part*, 100 F.3d 1150 (4th Cir. 1996), *cert. denied*, 522 U.S. 1044 (1998); Casey v. Ohio Med. Prods., 877 F. Supp. 1380, 1382 (N.D. Cal. 1995). For a discussion of specific causation, see *infra* § VII.

7. Richard Doll & A. Bradford Hill, *Lung Cancer and Other Causes of Death in Relation to Smoking*, 2 Brit. Med. J. 1071 (1956).

8. *See* Kelley v. American Heyer-Schulte Corp., 957 F. Supp 873, 878 (W.D. Tex. 1997), *appeal dismissed*, 139 F.3d 899 (5th Cir. 1998). Association is more fully discussed *infra* § 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. 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* § V.

9. *See infra* § IV.

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 most studies have flaws.[10] Some flaws are inevitable given the limits of technology and resources. In evaluating epidemiologic evidence, the key questions, then, are the extent to which a study's flaws compromise its findings and whether the effect of the flaws can be assessed and taken into account in making inferences.

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 these cases are discussed later in this reference guide.[11]

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 couple of decades, courts frequently have confronted the use of epidemiologic studies as evidence and 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 epidemiological data is generally a fit subject for judicial notice; epidemiology is a well-established branch of science and medicine, and epidemiological evidence has been accepted in numerous cases."[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 reveal an association between an agent and disease?
2. What sources of error in the study may have contributed to an inaccurate result?
3. If the agent is associated with disease, is the relationship causal?

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 er-

10. *See 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.").

11. *See infra* § VII.

12. 911 F.2d 941, 954 (3d Cir. 1990); *see also* Smith v. Ortho Pharm. Corp., 770 F. Supp. 1561, 1571 (N.D. Ga. 1991) (explaining increased reliance of courts on epidemiologic evidence in toxic substances litigation).

ror.[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 examines issues of specific causation, considering whether an agent caused an individual's disease.

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 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 disease or health outcome. 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 between the two groups in outcome 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 agent or treatment is given a placebo, and that use double blinding, 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 assessing the value of agents that are thought to be beneficial to human beings.

13. For a more in-depth discussion of the statistical basis of epidemiology, see David H. Kaye & David A. Freedman, Reference Guide on Statistics § 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.

When an agent's effects are suspected to be harmful, we cannot knowingly expose people to the agent.[14] Instead of the investigator controlling who is exposed to the agent and who is not, most epidemiologic studies are observational—that is, they "observe" a group of individuals who have been exposed to an agent of interest, such as cigarette smoking or an industrial chemical, and compare them with another group of individuals who have not been so exposed. Thus, the investigator identifies a group of subjects who have been knowingly or unknowingly exposed 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 contribute to the risk of developing the disease in question. Since 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).

B. The Types of Observational Study Design

Several different types of observational epidemiologic studies can be conducted.[15] Study designs may be chosen because of suitability for investigating the question of interest, timing constraints, resource limitations, or other considerations. An important question that might be asked initially about a given epidemiologic study is whether the study design used was appropriate to the research question.

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.[16] A final type of observational study, one in which data about individuals is not gathered, but rather population data about expo-

14. Experimental studies in which human beings are exposed to agents known or thought to be toxic are ethically proscribed. *See* Ethyl Corp. v. United States Envtl. Protection Agency, 541 F.2d 1, 26 (D.C. Cir.), *cert. denied*, 426 U.S. 941 (1976). 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.*, E.R. Squibb & Sons, Inc. v. Stuart Pharms., No. 90-1178, 1990 U.S. Dist. LEXIS 15788 (D.N.J. Oct. 16, 1990); Gordon H. Guyatt, *Using Randomized Trials in Pharmacoepidemiology*, *in* Drug Epidemiology and Post-Marketing Surveillance 59 (Brian L. Strom & Giampaolo Velo eds., 1992). Experimental studies may also be conducted that entail discontinuation of exposure to a harmful agent, such as studies in which smokers are randomly assigned to a variety of smoking-cessation programs or no cessation.

15. 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* § II.B.4.

16. *See infra* § II.B.3.

sure and disease are used, is an ecological study.

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"). Thus, a cohort study takes the exposed status of participants (the independent variable) and examines its effect on incidence of disease (the dependent variable). A case-control study takes the disease status as the independent variable and examines its relationship with exposure, which is the dependent variable. 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[17] the researcher identifies two groups of individuals: (1) individuals who have been exposed to a substance that is considered a possible cause of a disease and (2) individuals who have not been exposed (see Figure 1).[18] Both groups are followed for a specified length of time, and the proportions of individuals in each group who develop the disease are compared.[19] Thus, as illustrated in Table 1, a researcher would compare the proportion of unexposed individuals (controls) with the disease (b/(a + b)) with the proportion of exposed individuals (cohort) with the disease (d/(c + d)). If the exposure causes

17. Cohort studies also are referred to as prospective studies and follow-up studies.

18. 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 (1–2 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).

19. Sometimes retrospective cohort studies are conducted, in which the researcher gathers historical data about exposure and disease outcome of the exposed 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 records. Irving J. Selikoff et al., *The Occurrence of Asbestosis Among Insulation Workers in the United States,* 132 Annals N.Y. Acad. Sci. 139, 143 (1965).

the disease, the researcher would expect a greater proportion of the exposed individuals than of the unexposed individuals to develop the disease.[20]

Figure 1. Design of a Cohort Study

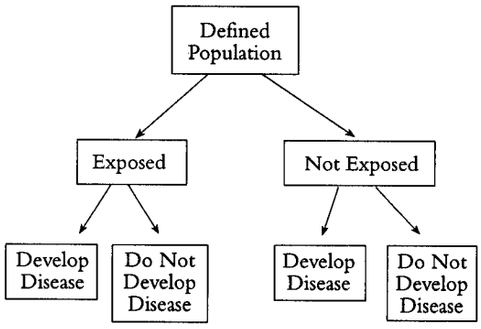


Table 1. Cross-Tabulation of Exposure by Disease Status

	No Disease	Disease
Not Exposed	a	b
Exposed	c	d

One advantage of the cohort study design is that the temporal relationship between exposure and disease can often be established more readily. By tracking the exposed and unexposed groups over time, the researcher can determine the time of disease onset. This temporal relationship is critical to the question of causation, since 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 compared with nonminers. The study group (also referred to as the exposed cohort) consisted of 3,400 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 ev-

20. 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 within a time period. All smokers and nonsmokers will, if followed for 100 years, die. Smokers will die at a greater rate than nonsmokers.

ery three 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.[21]

The cohort design is often used in occupational studies such as the one just cited. Since 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, it may be 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[22] 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,[23] the researcher begins with a group of individuals who have a disease (cases) and then selects a group of individuals who do not have the disease (controls). 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).

Thus, for example, in the late 1960s, doctors in Boston were confronted with an unusual incidence of vaginal adenocarcinoma in young female patients. 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

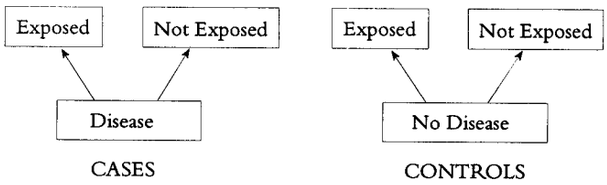
21. This example is based on a study description in Abraham M. Lilienfeld & David E. Lilienfeld, Foundations of Epidemiology 237–39 (2d ed. 1980). The original study is Joseph K. Wagoner et al., *Radiation as the Cause of Lung Cancer Among Uranium Miners*, 273 New Eng. J. Med. 181 (1965).

22. *See* Daniel L. Rubinfeld, Reference Guide on Multiple Regression § II.B, in this manual.

23. 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.

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.[24]

Figure 2. Design of a Case-Control Study



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.[25] 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, since 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 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),

24. *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).

25. 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 3,100 each for a cohort study, but only 177 each for a case-control study. Harold A. Kahn & Christopher T. Sempos, Statistical Methods in Epidemiology 66 (1989).

but these studies can provide valuable leads to further directions for research.[26]

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.[27] In contrast, studies that collect data only about the group as a whole are called ecological studies.[28] 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.[29] Such studies may be useful for identifying associations, but they rarely provide definitive causal answers. The difficulty is illustrated below with an ecological study of the relationship between dietary fat and cancer.

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).[30] 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

26. For more information (and references) about cross-sectional studies, see Leon Gordis, Epidemiology 137–39 (1996).

27. 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.

28. 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.

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

30. For a discussion of the data on this question and what they might mean, see David Freedman et al., Statistics (3d ed. 1998).

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 a demographic 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.[31] 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 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 animal studies (in vivo) may be used to determine toxicity in humans.[32] 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,[33] which epidemiology often confronts. Exposure can be carefully controlled and measured. Refusals to participate in a study are not an issue, and loss to follow-up 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

31. 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.

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

33. *See infra* § IV.C.

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.[34] 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.[35] 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.[36] Those matters are almost always fraught with considerable, and currently unresolvable, uncertainty.[37]

Toxicologists also use in vitro methods, in which human or animal tissue or cells are grown in laboratories and 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.[38]

Often toxicologic studies are the only or best available evidence of toxicity. Epidemiologic studies are difficult, time-consuming, and expensive, and consequently they do not exist for a large array of environmental agents. Where both animal toxicology and epidemiologic studies are available, no universal rules exist for how to interpret or reconcile them.[39] Careful assessment of the meth-

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

35. *See* Ian C.T. Nesbit & Nathan J. Karch, Chemical Hazards to Human Reproduction 98–106 (1983); International Agency for Research on Cancer (IARC), Interpretation of Negative Epidemiological Evidence for Carcinogenicity (N.J. Wald & R. Doll eds., 1985).

36. *See infra* § V.C & note 119.

37. *See* General Elec. Co. v. Joiner, 522 U.S. 136, 143–45 (1997) (holding that the district court did not abuse its discretion in excluding expert testimony on causation based on expert's failure to explain how animal studies supported expert's opinion that agent caused disease in humans).

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

39. *See* IARC, *supra* note 35 (identifying a number of substances and comparing animal toxicology evidence with epidemiologic evidence).

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), *cert. denied*, 494 U.S. 1046 (1990), 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), *cert. denied*, 493 U.S. 882 (1989); *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.

odological validity and power[40] 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.[41]

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 methods for investigating questions of causation—for example, toxicology and animal studies, clinical research, and epidemiology—which all have distinct advantages and disadvantages." *See also* Villari v. Terminix Int'l, Inc., 692 F. Supp. 568, 571 (E.D. Pa. 1988); Peterson v. Sealed Air Corp., Nos. 86-C3498, 88-C9859 Consol., 1991 U.S. Dist. LEXIS 5333, at *27–*29 (N.D. Ill. Apr. 23, 1991); *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), *cert. denied*, 499 U.S. 961 (1991). 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), *cert. denied*, 513 U.S. 1190 (1995); *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), *cert. denied*, 522 U.S. 1044 (1998).

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.), *cert. denied*, 506 U.S. 826 (1992); *In re* Paoli R.R. Yard PCB Litig., No. 86-2229, 1992 U.S. Dist. LEXIS 16287, at *16 (E.D. Pa. Oct. 21, 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 35; Bernard D. Goldstein & Mary Sue Henifin, Reference Guide on Toxicology § 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 not a 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."

40. *See infra* § IV.A.3.

41. *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.[42] 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 due to sampling error.

The strength of an association between exposure and disease can be stated as a relative risk, an odds ratio, or an attributable risk (often abbreviated as "RR," "OR," and "AR," respectively). 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:

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

The incidence rate of disease reflects 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.[43] Thus, the incidence rate expresses the risk that a

42. 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.

43. 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 26, at 32–34.

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 one 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.[44]

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. 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. Researchers should scrutinize their results for error. Error in the design of a study could yield an incorrect relative risk. Sources of bias and confounding should be examined.[45] Whenever an association is uncovered, further analysis should be conducted to determine if the association is real or due to an error or bias. Similarly, a study that does not find an association between an agent and disease may be erroneous because of bias or random error.

44. *See* DeLuca v. Merrell Dow Pharms., Inc., 911 F.2d 941, 947 (3d Cir. 1990); Gaul v. United States, 582 F. Supp. 1122, 1125 n.9 (D. Del. 1984).

45. *See infra* § IV.B–C.

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.[46] 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. The odds ratio approximates the relative risk when the disease is rare.[47]

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

Table 2. Cross-Tabulation of Cases and Controls by Exposure Status

	Cases	Controls
Exposed	a	b
Not Exposed	c	d

In a case-control study

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

Looking at the above 2 x 2 table, this ratio can be calculated as

$$\frac{a/c}{b/d}$$

This works out to ad/bc. Since 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 x 2 table (Table 3):

46. 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.

47. *See* Marcello Pagano & Kimberlee Gauvreau, Principles of Biostatistics 320–22 (1993). For further detail about the odds ratio and its calculation, see Kahn & Sempos, *supra* note 25, at 47–56.

Table 3. Case-Control Study Outcome

	Cases (with disease)	Controls (no disease)
Exposed	40	20
Not Exposed	60	80
Total	100	100

The calculation of the odds ratio would be

$$OR = \frac{40/60}{20/80} = 2.67$$

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.[48]

C. *Attributable Risk*

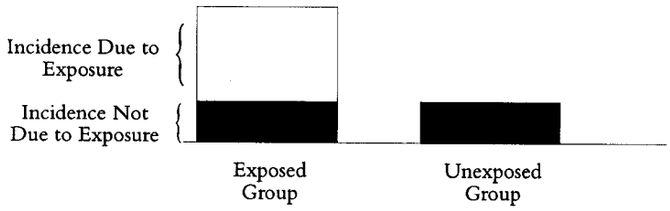
A frequently used measurement of risk is the attributable risk (AR). The attributable risk represents the amount of disease among exposed individuals that can be attributed to the exposure. It can also be expressed as the proportion of the disease among exposed individuals that is associated with the exposure (also called the "attributable proportion of risk," the "etiologic fraction" or "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.[49] 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).[50]

48. 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.

49. Kenneth J. Rothman & Sander Greenland, Modern Epidemiology 53–55 (2d ed. 1998). *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%, i.e., (1.55 - 1.0)/1.55 = .35 or 35%).

50. Risk is not zero for the control group (those not exposed) when there are other causal chains that cause the disease which 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. United States Envtl. Protection Agency, 541 F.2d 1, 25 (D.C. Cir.), *cert. denied*, 426 U.S. 941 (1976). There are some diseases that do not occur without exposure to an agent; these are known as signature diseases. *See infra* note 128.

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

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

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 one 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. Florida has a much higher death rate than Alaska.[51] Is sunshine dangerous? Perhaps, but the Florida population is much older than the Alaska population, and some adjustment must be made for the different age demo-

51. *See* Lilienfeld & Stolley, *supra* note 29, at 68–70 (mortality rate in Florida approximately three times what it is in Alaska).

graphics. The technique used to accomplish this is called adjustment, and two types of adjustment are used—direct and indirect.

In direct age adjustment, a standard population is used in order to eliminate the effects of any age differences between two study populations. Thus, in comparing two populations, A and B, the age-specific mortality rates for Population A are applied to each age group of the standard reference population, and the numbers of deaths expected in each age group of the standard population are calculated. These expected numbers of deaths are then totaled to yield the number of deaths expected in the standard population if it experienced the mortality risk of Population A. The same procedure is then carried out for Population B. Using these expected numbers of deaths, mortality rates are calculated for the standard population on the basis of the number of deaths expected if it had the mortality experience of Population A and the number of deaths expected if it had the mortality experience of Population B. We can then compare these 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.

A second approach, indirect age adjustment, is often used, for example, in studying mortality in an occupationally exposed population, such as miners or construction workers. To answer the question whether a population of miners has a higher mortality rate than we would expect in a similar population not engaged in mining, we must apply the age-specific rates for a known population, such as all men of the same age, to each age group in the population of interest. This will yield the number of deaths expected in each age group in the population of interest if this population had had the mortality experience of the known population. The number of deaths expected is thus calculated for each age group and totaled; the numbers of deaths that were actually observed in that population are counted. The ratio of the total number of deaths actually observed to the total number of deaths that would be expected if the population of interest actually had the mortality experience of the known population is then calculated. 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.[52] If the ratio equals 1.0, the observed number of deaths equals the expected number of deaths, and the mortality experience of the population of interest is no different from that of the known population. If the SMR is greater than 1.0, the population of interest has a higher mortality risk than that of the known population, and if the SMR is less than 1.0, the population of interest has a lower mortality risk than that of the known population.

52. *See In re* Joint E. & S. Dist. Asbestos Litig., 52 F.3d 1124, 1128 (2d Cir. 1995) (using SMR to describe relative risk of an agent in causing disease). 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).

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.[53]

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 association. Or a study may erroneously conclude 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.

There are three explanations why an association found in a study may be erroneous: chance, bias, and confounding. Before any inferences about causation are drawn from a study, the possibility of these phenomena must be examined.[54]

The findings of a study may be the result of chance (or sampling error) because virtually all epidemiologic studies are based on sampling a small proportion of the relevant population. During the design of a study, the size of the sample can be increased to reduce (but not eliminate) the likelihood of sampling error. Once a study has been completed, statistical methods (discussed in the next subsection) permit an assessment of whether the results of a study are likely to represent a true association or 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 the level of significance used entails a somewhat arbitrary determination.[55] A confidence interval

53. For further elaboration on adjustment, see Rothman & Greenland, *supra* note 49, at 234–35; Gordis, *supra* note 26, at 49–52; Philip Cole, *Causality in Epidemiology, Health Policy, and Law*, [1997] 27 Envtl. L. Rep. (Envtl. L. Inst.) 10279, 10281 (June 1997).

54. *See* Cole, *supra* note 53, at 10285. 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 (i.e., 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 § IV, in this manual.

55. Describing a study result as "statistically significant" does not mean that the result—the relative risk—is of a significant or substantial magnitude. *Statistical significance does not address the magnitude of the*

provides both the relative risk found in the study and a range (interval) within which the true relative risk resides with some (arbitrarily chosen) level of confidence. Both of these techniques are explained in subsection IV.A.

Bias (or systematic error) also can produce error in the outcome of a study. Epidemiologists attempt to minimize the existence of bias through their study design, which is developed before they begin gathering data. However, even the best designed and conducted studies can have biases, which may be subtle. Consequently, after a study is completed it should be evaluated for 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 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 analytic methods used.[56] Common biases and how they may produce invalid results are described in subsection 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 factor confounds the relationship being examined in the study. Confounding is explained in subsection IV.C.

A. What Statistical Methods Exist to Evaluate the Possibility of Sampling Error?[57]

Before detailing the statistical methods used to assess random error (which we use as synonymous with sampling error), we explain two concepts 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.

relative risk found in a study, only the likelihood that it would have resulted from random error if there is no real association between the agent and disease.

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

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

Scientists, including epidemiologists, generally begin an empirical study with a hypothesis that they seek to disprove,[58] called the null hypothesis. The null hypothesis states that there is no true association between an agent and a disease. Thus, the epidemiologist begins by technically assuming that the relative risk is 1.0 and seeks to develop data that may disprove the hypothesis.[59]

1. False positive error 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.[60] Random error is illustrated by a fair coin yielding five heads out of five tosses,[61] an occurrence that would result, purely by chance, in about 3% of a series of five tosses. Thus, even though the true relative risk is 1.0, an epidemiologic study may find a relative risk greater than 1.0 because of random error. 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 or 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 coin is tossed 1,000 times is more helpful than a test in which the coin is tossed only 10 times. Common sense dictates that it is far more likely that a test of a fair coin with 10 tosses will come up, for example, with

58. *See, e.g.,* Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579, 593 (1993) (scientific methodology involves generating and testing hypotheses). We should explain that this null-hypothesis testing model may be misleading. The reality is that the vast majority of epidemiologic studies are conducted because the researcher suspects that there is a causal effect and seeks to demonstrate that causal relationship. Nevertheless, epidemiologists prepare their study designs and test the plausibility that any association found in a study was the result of sampling error by using the null hypothesis.

59. *See* DeLuca v. Merrell Dow Pharms., Inc., 911 F.2d 941, 945 (3d Cir. 1990); Stephen E. Fienberg et al., *Understanding and Evaluating Statistical Evidence in Litigation*, 36 Jurimetrics J. 1, 21–24 (1995).

60. 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* Expert Evidence: A Practitioner's Guide to Law, Science, and the FJC Manual 91 (Bert Black & Patrick W. Lee eds., 1997).

61. *DeLuca,* 911 F.2d at 946–47.

80% heads than will a test with 1,000 tosses. For if the test is conducted with larger numbers (1,000 tosses), the stability of 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.[62]

One means for evaluating the possibility that an observed association could have occurred as a result of random error is by calculating a *p*-value.[63] A *p*-value represents the probability that a positive association would result from random error if no association were in fact present.[64] Thus, a *p*-value of .1 means that there is a 10% chance that if the true relative risk is 1.0, the observed relative risk (greater than 1.0) in the study was due to random error.[65]

To minimize false positive error, 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.[66] Thus, an outcome is statistically significant when the observed *p*-value for the study falls below the preselected significance level. The most common significance level, or alpha,

62. This explanation of numerical stability was drawn from Brief Amicus Curiae of Professor Alvan R. Feinstein in Support of Respondent 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), *cert. denied*, 484 U.S. 1004 (1988). 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."

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

64. Technically, a *p*-value represents the probability that the study's association or a larger one would occur as a result of sampling error where no association (or, equivalently, the null hypothesis) is the true situation. This means that if one conducted an examination of 20 associations in which the true RR = 1, on average one of those examinations would result in a statistically significant, yet spurious, association.

Unfortunately, some have failed to appreciate the difference between a statement of the probability that the study's outcome would occur as a result of random error (the correct understanding of a *p*-value) if the true association were RR equal to 1 and a statement of the probability that the study's outcome was due to random error (an incorrect understanding of a *p*-value). *See, e.g.*, *In re* TMI Cases Consol. II, 922 F. Supp. 997, 1017 (M.D. Pa. 1996); Barnes v. Secretary of Dep't of Health & Human Servs., No. 92-0032V, 1997 U.S. Claims LEXIS 212, at *22 (Fed. Cl. Sept. 15, 1997) ("The *P* value . . . [measures] the probability that the results could have happened by chance alone."). Conventional statistical methodology does not permit calculation of the latter probability. However, the *p*-value is used to assess the plausibility that a positive association should be taken to disprove the null hypothesis and permit an inference, after assessing the factors discussed in section V *infra*, that the agent causes disease.

65. 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.

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

used in science is .05.[67] 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.[68] Although .05 is often the significance level selected, other levels can and have been used.[69] Thus, in its study of the effects of secondhand smoke, the Environmental Protection Agency (EPA) used a .10 standard for significance testing.[70]

67. 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.,* Ethyl Corp. v. United States Envtl. Protection Agency, 541 F.2d 1, 28 n.58 (D.C. Cir.), *cert. denied,* 426 U.S. 941 (1976); Hodges v. Secretary of Dep't of Health & Human Servs., 9 F.3d 958, 967, 970 (Fed. Cir. 1993) (Newman, J., dissenting); Edward J. Imwinkelried, *The Admissibility of Expert Testimony in* Christophersen v. Allied-Signal Corp.*: The Neglected Issue of the Validity of Nonscientific Reasoning by Scientific Witnesses,* 70 Denv. U. L. Rev. 473, 478 (1993).

This claim is incorrect, 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* § 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. *See* DeLuca v. Merrell Dow Pharms., Inc., 911 F.2d 941, 947 (3d Cir. 1990). 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 39, 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 § IV.B.2, in this manual; *Developments in the Law, supra* note 66, at 1551–56; Allen v. United States, 588 F. Supp. 247, 417 (D. Utah 1984) ("Whether a correlation between a cause and a group of effects is more likely than not—particularly in a legal sense—is a different question from that answered by tests of statistical significance . . . ."), *rev'd on other grounds,* 816 F.2d 1417 (10th Cir. 1987), *cert. denied,* 484 U.S. 1004 (1988); Turpin v. Merrell Dow Pharms., Inc., 959 F.2d 1349, 1357 n.2 (6th Cir.), *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.").

68. 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 associations, 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 Science 1538, 1539 (1994).

69. A significance test can be either one-tailed or two-tailed, depending on the null hypothesis selected by the researcher. Since 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. For an explanation of the difference between one-tailed and two-tailed tests, see David H. Kaye & David A. Freedman, Reference Guide on Statistics § IV.C.2, in this manual.

70. U.S. Envtl. 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

Statistical significance is a term that speaks only to the question of sampling error—it does not address the magnitude of any association found in a study.[71] A study may be statistically significant but may find only a very weak association; conversely, a study with small sample sizes may find a high relative risk but still not be statistically significant.[72]

There is some controversy among epidemiologists and biostatisticians about the appropriate role of significance testing.[73] To the strictest significance testers, 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

95%, though 90% (which corresponds to an alpha of .10) is also used; selection of the value is "somewhat arbitrary").

71. Unfortunately, some courts have been confused about the relationship between statistical significance and the magnitude of the association. *See In re* Joint E. & S. Dist. Asbestos Litig., 827 F. Supp. 1014, 1041 (S.D.N.Y. 1993), *rev'd on other grounds,* 52 F.3d 1124 (2d Cir. 1995) (concluding that any relative risk less than 1.50 is statistically insignificant).

72. *See* Cole, *supra* note 53, at 10282. While statistical significance and association are two distinct concepts, whether a study's results are statistically significant does depend, in part, on the incidence of disease and the magnitude of any association found in the study. In other words, the more common the disease and the greater the association between an agent and the disease, the more likely that a study's outcome will be statistically significant, all other things being equal. Also critical to alpha is the number of persons participating in the study. As the disease becomes more infrequent, the sample sizes decrease, and the associations found are weaker, it is less likely that the results will be statistically significant.

73. 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.), *amended,* 884 F.2d 167 (5th Cir. 1989), *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 v. Merrell Dow Pharms., Inc., 884 F.2d 167, 167 (5th Cir. 1989).

A number of courts have followed the *Brock* decision or have indicated strong support for significance testing as a screening device. *See* Kelley v. American Heyer-Schulte Corp., 957 F. Supp. 873, 878 (W.D. Tex. 1997) (lower end of 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); Thomas v. Hoffman-LaRoche, Inc., 731 F. Supp. 224, 228 (N.D. Miss. 1989) (granting judgment n.o.v. and observing that "there is a total absence of any statistically significant study to assist the jury in its determination of the issue of causation"), *aff'd on other grounds,* 949 F.2d 806 (5th Cir.), *cert. denied,* 504 U.S. 956 (1992); Daubert v. Merrell Dow Pharms., Inc., 727 F. Supp. 570, 575 (S.D. Cal. 1989), *aff'd on other grounds,* 951 F.2d 1128 (9th Cir. 1991), *vacated,* 509 U.S. 579 (1993); Wade-Greaux v. Whitehall Labs., Inc., 874 F. Supp. 1441 (D.V.I. 1994); Merrell Dow Pharms., Inc. v. Havner, 953 S.W.2d 706, 724 (Tex. 1997).

By contrast, a number of courts appear more cautious about 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 within which the true value is thought to lie, with a specified level of confidence)

critical of using strict significance testing, which rejects all studies with an observed *p*-value below that specified level. Epidemiologic studies have become increasingly sophisticated in addressing the issue of random error and examining the data from studies to ascertain what information they may provide about the relationship between an agent and a disease, without the rejection of all studies that are not statistically significant.[74]

Calculation of a confidence interval permits a more refined assessment of appropriate inferences about the association found in an epidemiologic study.[75] A confidence interval is 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. The advantage of a confidence interval is that it displays more information than significance testing. What a statement about whether a result is statistically significant does not provide is the magnitude of the association found in the study or an indication of how statistically stable that association is. A confidence interval for any study shows the relative risk determined in the study as a point on a numerical axis. It also displays the boundaries of relative risk

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* Turpin v. Merrell Dow Pharms., Inc., 959 F.2d 1349, 1357 (6th Cir.) ("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."), *cert. denied*, 506 U.S. 826 (1992); *In re* Bendectin Prod. Liab. Litig., 732 F. Supp. 744, 748–49 (E.D. Mich. 1990) (rejecting defendant's claim that plaintiff could not prevail without statistically significant epidemiologic evidence); Berry v. CSX Transp., Inc., 709 So. 2d 552, 570 (Fla. Dist. Ct. App. 1998) (refusing to hold studies that were not statistically significant inadmissible).

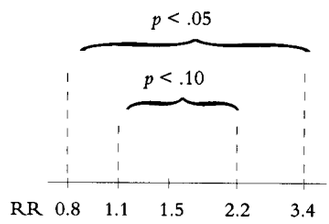
Although the trial court had relied in part on the absence of statistically significant epidemiologic studies, the Supreme Court in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993), did not explicitly address the matter. The Court did, however, refer to "the known or potential rate 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 66, 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.

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

75. 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 confidence intervals. In *Turpin*, 959 F.2d at 1353–54 n.1, the court discussed the relationship among confidence intervals, alpha, and power. 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 § IV.A, in this manual.

consistent with the data found in the study based on one or several selected levels of alpha or statistical significance. An example of two confidence intervals that might be calculated for a study is displayed in Figure 4.

Figure 4. Confidence Intervals



The confidence interval shown in Figure 4 represents a study that found a relative risk of 1.5, with boundaries of 0.8 to 3.4 for alpha equal to .05 (equivalently, a confidence level of .95) and boundaries of 1.1 to 2.2 for alpha equal to .10 (equivalently, a confidence level of .90). Because the boundaries of the confidence interval with alpha set at .05 encompass a relative risk of 1.0, the study is not statistically significant at that level. By contrast, since the confidence boundaries for alpha equal to .10 do not include a relative risk of 1.0, the study does have a positive finding that is statistically significant at that level of alpha. The larger the sample size in a study (all other things being equal), the narrower the confidence boundaries will be (indicating greater statistical stability), thereby reflecting the decreased likelihood that the association found in the study would occur if the true association is 1.0.[76]

76. Where multiple epidemiologic studies are available, a technique known as meta-analysis (*see infra* § 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 pooling 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. Gerbarg & Ralph I. Horwitz, *Resolving Conflicting Clinical Trials: Guidelines for Meta-Analysis*, 41 J. Clinical Epidemiology 503 (1988).

In *In re Paoli Railroad Yard PCB Litigation*, 916 F.2d 829, 856–57 (3d Cir. 1990), *cert. denied*, 499 U.S. 461 (1991), 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 (FDA) pooled data from

2. False negative error

False positives can be reduced by adopting more stringent values for alpha. Using a level of .01 or .001 will result in fewer false positives than using an alpha of .05. The trade-off 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 not to disprove the null hypothesis when in fact there is a true association of a specified magnitude.[77] 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;[78] that is, beta, or the likelihood of erroneously failing to reject the null hypothesis, depends on the selection of an alternative hypothesis about the magnitude of association and the level of alpha chosen.

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. The concept of power can be helpful in evaluating whether a study's outcome is exonerative or inconclusive.[79]

The power of a study expresses 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.[80] 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.[81]

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[82] 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

control groups in different studies in which some gave the controls a placebo and others gave the controls an alternative treatment), *cert. denied*, 510 U.S. 914 (1993).

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

78. *See* Green, *supra* note 39, at 684–89.

79. *See* Fienberg et al., *supra* note 59, at 22–23.

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

81. For examples of power curves, see Kenneth J. Rothman, Modern Epidemiology 80 (1986); Pagano & Gauvreau, *supra* note 47, at 223.

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

is low, it has significantly less probative value than a study with similar results but a higher power.[83]

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

Systematic error or bias can produce an erroneous association in an epidemiologic study. Bias may arise in the design or conduct of a study, data collection, or data analysis. When scientists use the term *bias,* it does not necessarily carry an imputation of prejudice or other subjective factors, such as the researcher's desire for a particular outcome. The meaning of scientific bias differs from conventional (and legal) usage, in which bias refers to a partisan point of view.[84] Bias refers to anything (other than random sampling error) that results in error in a study and thereby compromises its validity. The two main classes of bias are selection bias (inappropriate 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 is due to the method of selection of cases and controls (in a case-control study) or exposed and unexposed individuals (in a cohort study).[85] The selection of an appropriate

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

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

85. 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 84, 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), *cert. denied,* 484 U.S. 1004 (1988), 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 due to 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.

control group has been described as the Achilles' heel of a case-control study.[86] Selecting members of the control group (those without disease) is problematic in case-control studies if the control participants were selected for reasons that are related to their having the exposure or potential risk factor being studied.

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 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.[87] 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 case-control study controls should be selected independently of their exposure status, in cohort studies, unexposed controls should be selected independently of their disease risk. For example, in a cohort study of cervical cancer, those who are not at risk for the disease—women who have had their cervices removed and men—should be excluded from the study population. Inclusion of such individuals as controls in a cohort study could result in erroneous findings by overstating the association between the agent and the disease.

A further source of selection bias occurs when those selected to participate refuse 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 refuses to participate in the study, the researcher should investigate reasons for refusal and whether those who refused are different from those who agreed. The researcher can show that those in the study are not a biased sample by comparing relevant characteristics of individuals who refused to participate with those of individuals who participated to show the similarity of the groups or the degree of differences. Similarly, if a significant number of subjects drop out of a study before completion, there may be a problem in determining whether the remaining subjects are representative of the original study populations. The researcher should

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

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

examine whether the study groups are still representative of the original study populations.

The fact that a study may suffer from selection bias does not in itself 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, in studies with multiple control groups, the consistent finding of an association when cases are compared with different control groups suggests that possible biases applicable to a particular control group are not invalidating.

2. Information bias

Information bias refers to the bias resulting from inaccurate information about the study participants regarding either their disease or exposure status. 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. In some situations the researcher is required to interview the subjects about past exposures, thus relying on the subjects' memories. Research has shown that individuals with disease (cases) may more readily recall past exposures than individuals with no disease (controls);[88] 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 malformations were caused by an infection during the mother's pregnancy.[89] A group of mothers of malformed infants (cases) and a group of mothers of infants with no 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. The issue of recall bias can sometimes be evaluated by finding a second source of data to validate the subject's response

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

89. *See* Brock v. Merrell Dow Pharms., Inc., 874 F.2d 307, 311–12 (5th Cir. 1989) (discussion of recall bias among women who bear children with birth defects), *cert. denied*, 494 U.S. 1046 (1990). 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 random 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 39, 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).

(e.g., blood test results from prenatal visits or medical records that document symptoms of infection).[90] 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, 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.

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.[91] 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 are only available for a small number of toxins and only reveal whether a person was exposed. Biological markers rarely help determine the intensity or duration of exposure.[92]

Monitoring devices also can be used to measure exposure directly but often are not available for exposures that occurred in the past. For past exposures, epidemiologists often use indirect means of measuring 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. If the agent of interest is a drug, medical or hospital records can be used to determine past exposure. Thus, retrospective

90. 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).

91. *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).

92. Dose generally refers to the intensity or magnitude of exposure multiplied by the time exposed. *See* Sparks v. Owens-Illinois, Inc., 38 Cal. Rptr. 2d 739, 742 (Ct. App. 1995). 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), *cert. denied*, 484 U.S. 1004 (1988). 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 only likely to cause birth defects 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 simi-

occupational or environmental measurements of exposure are usually less accurate than prospective studies or follow-up studies, especially ones in which a drug or medical intervention is the independent variable being measured.

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.[93]

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.[94] 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, since employees tend to keep such information private. If the researcher uses an unreliable source of data, the study may not be useful to the court.

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

lar to those faced by an epidemiologist planning a study. Thus, in *Christophersen v. Allied-Signal Corp.*, 939 F.2d 1106, 1113 (5th Cir. 1991), *cert. denied*, 503 U.S. 912 (1992), the court criticized the plaintiff's expert, who relied on an affidavit of a co-worker to determine the dose of nickel and cadmium to which the decedent had been exposed.

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).

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

94. 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).

disease should be assessed. 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.

The scientific validity of the research findings is influenced by the reliability of the diagnosis of disease or health status.[95] For example, a researcher interested in studying spontaneous abortion in the first trimester needs to test women for pregnancy. Diagnostic criteria that are accepted by the medical community should be used to make the diagnosis. If a diagnosis is made using an unreliable home pregnancy kit 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 form 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. Misclassification bias has been subdivided into differential misclassification and nondifferential misclassification. Nondifferential misclassification occurs when inaccuracies in determining exposure are independent of disease status or when inaccuracies in diagnoses are independent of exposure status. This is a common problem resulting from the limitations of data collection. 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 grounds that data were inaccurately classified. Instead, nondifferential misclassification generally serves to reduce the observed association below its true magnitude.

Differential misclassification refers to the differential 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 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 an association.

95. 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.

3. Other conceptual problems

Sometimes studies are flawed because of flawed definitions or premises that do not fall under the rubric of selection bias or information bias. For example, if the researcher defines the disease of interest as all birth defects, rather than a specific birth defect, he or she must have a scientific basis to hypothesize that the effects of the agent being investigated could be so varied. 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.[96]

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 must consider the possibility of confounding factors. The expert must also evaluate several criteria to determine whether an inference of causation is appropriate. These matters are discussed below.

C. Could a Confounding Factor Be Responsible for the Study Result?[97]

Even when an association exists, researchers must determine whether the exposure causes the disease or whether the exposure and disease are caused by some other confounding factor. A confounding factor 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.[98] Researchers must separate the relationship be-

96. In *Brock v. Merrell Dow Pharmaceuticals, Inc.*, 874 F.2d 307, 312 (5th Cir. 1989), *cert. denied*, 494 U.S. 1046 (1990), 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 & Greenland, *supra* note 49, at 132 ("Unwarranted assurances of a lack of any effect can easily emerge from studies in which a wide range of etiologically unrelated outcomes are grouped.").

97. *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).

98. This example is drawn from Kahn & Sempos, *supra* note 25, at 63.

tween 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.[99] Some 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. Failure to recognize confounding can introduce a bias—error—into the findings of the study.

In 1981, Dr. Brian MacMahon, Professor and Chairman of the Department of Epidemiology at the Harvard School of Public Health, reported an association between coffee drinking and cancer of the pancreas in the *New England Journal of Medicine*.[100] This observation caused a great stir, and in fact, one coffee distributor ran a large advertisement in the *New York Times* refuting the findings of the study. What could MacMahon's findings mean? One possibility is that the association is causal and that drinking coffee causes an increased risk of cancer of the pancreas. However, there is also another possibility. We know that smoking is an important risk factor for cancer of the pancreas. We also know that it is difficult to find a smoker who does not drink coffee. Thus, drinking coffee and smoking are associated. An observed association between coffee consumption and an increased risk of cancer of the pancreas could reflect the fact that smoking causes cancer of the pancreas and that smoking also is associated closely with coffee consumption. The association MacMahon found between drinking coffee and pancreatic cancer could be due to the confounding factor of smoking. To be fair to MacMahon, we must note that he was aware of the possibility of confounding and took it into account in his study design by gathering and analyzing data separately for smokers and nonsmokers. The association between coffee and pancreatic cancer remained even when smoking was taken into account.

The main problem in many observational studies such as MacMahon's is that the individuals are not assigned randomly to the groups being compared.[101] As discussed above, randomization maximizes the possibility that exposures other

99. 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 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.

100. Brian MacMahon et al., *Coffee and Cancer of the Pancreas*, 304 New Eng. J. Med. 630 (1981).

101. 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. For additional comments on randomization and confounding, see the Glossary of Terms.

than the one under study are evenly distributed between the exposed and the control cohorts.[102] 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 controls by residence, socioeconomic status, age, or other extraneous factors.[103] These other factors may be causing 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 MacMahon study, do not reflect an error made by the investigators; rather, they reflect the inherently "uncontrolled" nature of observational studies. When they can be identified, confounders should be taken into account. Confounding factors that are suspected or known in advance can be controlled during the study design through study-group selection. Unanticipated confounding factors that are suspected after data collection can sometimes be controlled during data analysis, if data have been gathered about them.

MacMahon's study found that coffee drinkers had a higher rate of pancreatic cancer than those who did not drink coffee. To evaluate whether smoking is a confounding factor, the researcher would divide each of the exposed and control groups into smoking and nonsmoking subgroups to examine whether subjects' smoking status affects the study results. If the outcome in the smoking subgroups is the same as that in the nonsmoking subgroups, smoking is not a confounding factor. If the subjects' smoking status affects the outcome, then smoking is a confounder, for which adjustment is required. If the association between coffee drinking and pancreatic cancer completely disappears when the subjects' smoking status is considered, then smoking is a confounder that fully accounts for the association with coffee observed. Table 4 reveals a hypothetical study's results, with smoking being a weak confounding factor, which, when accounted for, does not eliminate the association between coffee drinking and cancer.

102. *See* Rothman & Greenland, *supra* note 49, at 124; *see also supra* § II.A.

103. *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), *cert. denied*, 484 U.S. 1004 (1988).

Table 4. Pancreatic Cancer Study Data

Pancreatic Cancer Status	All Subjects		Smokers >1 Pack per Day		Nonsmokers	
	Controls	Coffee Drinkers	Controls	Coffee Drinkers	Controls	Coffee Drinkers
Cancer	14	17	8	11	6	6
No Cancer	1,393	476	733	263	660	213
RR	1.1	3.9	1.2	4.6	1.0	3.1

Note: RR = relative risk.

There is always a real risk that an undiscovered or unrecognized confounding factor may contribute to a study's findings, by either magnifying or reducing the observed association.[104] 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 those 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.

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. When a factor or factors, such as age, sex, or even smoking status, are considered 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.[105]

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.

104. Rothman & Greenland, *supra* note 49, at 120; *see also supra* § II.A.

105. 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.

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 case of MacMahon's study, the researcher would evaluate whether smoking is a confounding factor by comparing the risk of pancreatic cancer in all coffee drinkers (including smokers) with the risk in nonsmoking coffee drinkers. If the risk is substantially the same, smoking is not a confounding factor (e.g., smoking does not distort the relationship between coffee drinking and the development of pancreatic cancer), which is what MacMahon found. If the risk is substantially different, but still exists in the nonsmoking group, then smoking is a confounder, but doesn't wholly account for the association with coffee. If the association disappears, then smoking is a confounder that fully accounts for the association with coffee observed.

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

To control for confounding factors during data analysis, researchers can use one of two techniques: stratification or multivariate analysis.

Stratification reduces or eliminates confounding by evaluating the effect of an exposure at different levels (strata) of exposure to the confounding variable. Statistical methods then can be applied to combine the results of exposure at each stratum into an overall single estimate of risk. For example, in MacMahon's study of smoking and pancreatic cancer, if smoking had been a confounding factor, the researchers could have stratified the data by creating subgroups based on how many cigarettes each subject smoked a day (e.g., a nonsmoking group, a light smoking group, a medium smoking group, and a heavy smoking group). When different rates of pancreatic cancer for people in each group who drink the same amount of coffee are compared, the effect of smoking on pancreatic cancer is revealed. The effect of the confounding factor can then be removed from the study results.

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.[106]

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 then applies the guidelines described in section V to determine whether an inference of causation is warranted.

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

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 with, the way the familiar "but for," or sine qua non, test is used in law for cause in fact. "An act or an omission is not regarded as a cause of an event if the particular event would have occurred without it."[107] This is equivalent to describing the act or occurrence as a necessary link in a chain of events that results in the particular event.[108] 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. Thus, exposure is a necessary condition for the increase in the incidence of disease among those exposed.[109] 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 objectively prove causation; rather, causation is a judgment for epidemiologists and others interpreting the epidemiologic data. 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 were discussed in section IV. Once this process is completed, researchers consider how guidelines

107. W. Page Keeton et al., Prosser and Keeton on the Law of Torts 265 (5th ed. 1984); *see also* Restatement (Second) of Torts § 432(1) (1965).

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 cause. Neither fire is a but-for, or necessary condition, for the destruction of the house, because either fire would have destroyed the house. *See id.* § 432(2). 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. 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.

108. *See supra* note 8.

109. *See* Rothman & Greenland, *supra* note 49, 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), *cert. denied*, 484 U.S. 1004 (1988).

for inferring causation from an association apply to the available evidence. These guidelines consist of several key inquiries that assist researchers in making a judgment about causation.[110] Most 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.[111]

The factors that guide epidemiologists in making judgments about causation are

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. One or more factors may be absent even when a true causal relationship exists. Similarly, the existence of some factors does not ensure that a causal relationship exists. Drawing causal inferences after finding an association and considering these factors requires judgment and searching analysis, based on biology, of why a factor or factors may be absent despite a causal relationship, and vice-versa. While the drawing of causal inferences is informed by scientific expertise, it is not a determination that is made by using scientific methodology.

110. *See* Mervyn Susser, Causal Thinking in the Health Sciences: Concepts and Strategies in Epidemiology (1973); *In re* Joint E. & S. Dist. Asbestos Litig., 52 F.3d 1124, 1128–30 (2d Cir. 1995) (discussing lower courts' use of factors to decide whether an inference of causation is justified when an association exists).

111. 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.), *cert. denied*, 479 U.S. 950 (1986).

These guidelines reflect criteria proposed by the U.S. Surgeon General in 1964[112] in assessing the relationship between smoking and lung cancer and expanded upon by A. Bradford Hill in 1965.[113]

A. Is There a Temporal Relationship?

A temporal, or chronological, relationship must exist for causation. If an exposure causes disease, the exposure must occur before the disease develops.[114] If the exposure occurs after the disease develops, it cannot cause the disease. Although temporal relationship is often listed as one of many factors in assessing whether an inference of causation is justified, it is a necessary factor: Without exposure before disease, causation cannot exist.

B. How Strong Is the Association Between the Exposure and Disease?[115]

The relative risk is one of the cornerstones for causal inferences.[116] Relative risk measures the strength of the association. The higher the relative risk, the greater the likelihood that the relationship is causal.[117] For cigarette smoking, for example, the estimated relative risk for lung cancer is very high, about 10.[118] That is, the risk of lung cancer in smokers is approximately ten 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

112. U.S. Dep't of Health, Educ., and Welfare, Public Health Serv., Smoking and Health: Report of the Advisory Committee to the Surgeon General (1964).

113. A. 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*.").

114. *See* Carroll v. Litton Sys., Inc., No. B-C-88-253, 1990 U.S. Dist. LEXIS 16833, at *29 (W.D.N.C. Oct. 29, 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 . . . .").

115. 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* § VII.

116. *See supra* § III.A.

117. *See* Cook v. United States, 545 F. Supp. 306, 316 n.4 (N.D. Cal. 1982); 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 39, 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 & Greenland, *supra* note 49, at 9–11.

118. *See* Doll & Hill, *supra* note 7.

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.

C. Is There a Dose–Response Relationship?

A dose–response relationship means that the more intense the exposure, the greater the risk of disease. Generally, higher exposures should increase the incidence (or severity) of disease. 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).[119] Thus, a dose–response relationship is strong, but not essential, evidence that the relationship between an agent and disease is causal.

D. Have the Results Been Replicated?

Rarely, if ever, does a single study conclusively demonstrate a cause–effect relationship.[120] 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.

The need to replicate research findings permeates most fields of science. In epidemiology, research findings often are replicated in different populations.[121] Consistency in these findings is an important factor in making a judgment about causation. Different studies that examine the same exposure–disease relationship

119. 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 Annals 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 Annals N.Y. Acad. Sci. 205 (1979); Ferebee v. Chevron Chem. Co., 736 F.2d 1529, 1536 (D.C. Cir.) (dose–response relationship for low doses is "one of the most sharply contested questions currently being debated in the medical community"), *cert. denied*, 469 U.S. 1062 (1984); *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), *cert. denied*, 484 U.S. 1004 (1988); 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).

120. In *Kehm v. Procter & Gamble Co.*, 580 F. Supp. 890, 901 (N.D. Iowa 1982), *aff'd sub nom.* Kehm v. Procter & Gamble Mfg. Co., 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.

121. *See* Cadarian v. Merrell Dow Pharms., Inc., 745 F. Supp. 409, 412 (E.D. Mich. 1989) (hold-

generally should yield similar results. While inconsistent results do not 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)?[122]

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. When an observation is inconsistent with current biological knowledge, it should not be discarded, but 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 better than the mechanisms of others.

F. Have Alternative Explanations Been Considered?

The importance of considering the possibility of bias and confounding and ruling out the possibilities was discussed above.[123]

G. What Is the Effect of Ceasing Exposure?

If an agent is a cause of a disease 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.

ing 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").

122. A number of courts have adverted to this criterion in the course of their discussions of causation in toxic substances cases. *E.g.*, 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, § III.E, in this manual.

123. *See supra* § IV.B–C.

H. Does the Association Exhibit Specificity?

An association exhibits specificity if the exposure is associated only with a single disease or type of disease.[124] The vast majority of agents do not cause a wide variety 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, while evidence of specificity may strengthen the case for causation, lack of specificity does not necessarily undermine it where there is a plausible 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.

124. This criterion reflects the fact that although an agent causes one disease, it does not necessarily cause other diseases. *See, e.g.*, Nelson v. American 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. International Flavors & Fragrances, Inc., 950 F. Supp. 981, 996–98 (C.D. Cal. 1996).

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

Not infrequently, the court may be faced with 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 which report associations, but of different magnitude. In view of the fact that epidemiologic 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.[125] Meta-analysis is a method of pooling study results to arrive at a single figure to represent the totality of the studies reviewed. 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.[126]

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 non-randomized 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 problematic. 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 not always apparent.

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? How can the problem of differences in the quality of the studies reviewed be taken into account? Can the results of the meta-analysis itself be reproduced by other analysts? When there

125. *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. Consolidated Rail Corp., 926 F.2d 262, 273 (3d Cir. 1991); Allen v. International 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 that sampling error is biasing the outcome. *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.

126. Petitti, *supra* note 76.

are several meta-analyses of a given relationship, why do the results of different meta-analyses often disagree? Another consideration is that often the differences among the individual studies included in a meta-analysis and the reasons for the differences are important in themselves and need to be understood; however, they may be masked in a meta-analysis. A final problem with meta-analyses is that they generate a single estimate of risk and may lead to a false sense of security regarding the certainty of the estimate. 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 like epidemiologic ones, may consequently be overlooked.[127]

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

Epidemiology is concerned with the incidence of disease in populations and does not address the question of the cause of an individual's disease.[128] This question, sometimes referred to as specific causation, is beyond the domain of the science of epidemiology. Epidemiology has its limits at the point where an

127. 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, Bailar has written the following:

> [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).

128. *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."); Smith v. Ortho Pharm. Corp., 770 F. Supp. 1561, 1577 (N.D. Ga. 1991); Grassis v. Johns-Manville Corp., 591 A.2d 671, 675 (N.J. Super. Ct. App. Div. 1991); 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., Winter 1986, at 93, 101. Asbestosis is a signature disease for asbestos, and adenocarcinoma (in young adult women) is a signature disease for in utero DES exposure. *See In re* "Agent Orange" Prod. Liab. Litig., 597 F. Supp. 740, 834 (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), *cert. denied*, 484 U.S. 1004 (1988).

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, epidemiology addresses whether an agent can cause a disease, not whether an agent did cause a specific plaintiff's disease.[129]

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, a number of courts have confronted the legal question of what is acceptable proof of specific causation and the role that epidemiologic evidence plays in answering that question.[130] This question is not a question that is addressed by epidemiology.[131] Rather, it is a legal question a number of courts have grappled with. An explanation of how these courts have resolved this question follows. The remainder of this section should be understood as an explanation of judicial opinions, not as epidemiology.

Before proceeding, one last 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.[132]

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,[133] as it tends to make an issue in dispute more or less likely.[134]

129. *Cf. "Agent Orange,"* 597 F. Supp. at 780.

130. 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.

131. Nevertheless, an epidemiologist may be helpful to the fact finder 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), *cert. dismissed*, 520 U.S. 1205 (1997); 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); 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).

132. *See* Green, *supra* note 39, at 672–73; 2 Modern Scientific Evidence, *supra* note 2, § 28-1.3.2 to -1.3.3, at 306–11.

133. *See DeLuca*, 911 F.2d at 958; *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

Far more courts have confronted the role that epidemiology plays with regard to the sufficiency of the evidence and the burden of production. The civil burden of proof is described most often as requiring the fact finder to "believe that what is sought to be proved . . . is more likely true than not true."[135] 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.[136] 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

concluded that an association between tampon use and menstrually related TSS [toxic shock syndrome] cases exists."), *aff'd sub nom.* Kehm v. Procter & Gamble Mfg. Co., 724 F.2d 613 (8th Cir. 1984).

Hearsay concerns may limit the independent admissibility of the study (*see supra* note 3), 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 . . . . "), *cert. denied*, 503 U.S. 912 (1992).

134. 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 epidemiologic 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.

135. 2 Edward J. Devitt & Charles B. Blackmar, Federal Jury Practice and Instruction § 71.13 (3d ed. 1977); *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), *cert. denied*, 444 U.S. 1073 (1980).

136. An adherent of the frequentist school of statistics would resist this adaptation, which may explain why so 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*

it is recognized as a cause of that disease in general."[137] The following discussion should be read with this caveat in mind.[138]

The threshold for concluding that an agent was more likely than not the cause of an individual's disease is a relative risk greater than 2.0. Recall that a relative risk of 1.0 means that the agent has no effect on the incidence of disease. When the relative risk reaches 2.0, the agent is responsible for an equal number of cases of disease as all other background causes. Thus, a relative risk of 2.0 (with certain qualifications noted below) implies a 50% likelihood that an exposed individual's disease was caused by the agent. A relative risk greater than 2.0 would permit an inference that an individual plaintiff's disease was more likely than not caused by the implicated agent.[139] A substantial number of courts in a variety of toxic substances cases have accepted this reasoning.[140]

also Steve Gold, Note, *Causation in Toxic Torts: Burdens of Proof, Standards of Persuasion and Statistical Evidence*, 96 Yale L.J. 376, 382–92 (1986). Subjective probabilities about discrete events are the product of adherents to Bayes Theorem. *See* Kaye, *supra* note 67, at 54–62; David H. Kaye & David A. Freedman, Reference Guide on Statistics § IV.D, in this manual.

137. Cole, *supra* note 53, at 10284.

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

139. *See* Davies v. Datapoint Corp., No. 94-56-P-DMC, 1995 U.S. Dist. LEXIS 21739, at *32–*35 (D. Me. Oct. 31, 1995) (holding that epidemiologist could testify about specific causation, basing such testimony on the probabilities derived from epidemiologic evidence).

140. *See* DeLuca v. Merrell Dow Pharms., Inc., 911 F.2d 941, 958–59 (3d Cir. 1990) (Bendectin allegedly caused limb reduction birth defects); *In re* Joint E. & S. Dist. Asbestos Litig., 964 F.2d 92 (2d Cir. 1992) (relative risk less than 2.0 may still be sufficient to prove causation); Daubert v. Merrell Dow Pharms., Inc., 43 F.3d 1311, 1320 (9th Cir.) (requiring that plaintiff demonstrate a relative risk of 2), *cert. denied*, 516 U.S. 869 (1995); Pick v. American Med. Sys., Inc., 958 F. Supp. 1151, 1160 (E.D. La. 1997) (recognizing that a relative risk of 2 implies a 50% probability of specific causation, but recognizing that a study with a lower relative risk is admissible, although ultimately it may be insufficient to support a verdict on causation); Sanderson v. International Flavors & Fragrances, Inc., 950 F. Supp. 981, 1000 (C.D. Cal. 1996) (acknowledging a relative risk of 2 as a threshold for plaintiff to prove specific causation); Manko v. United States, 636 F. Supp. 1419, 1434 (W.D. Mo. 1986) (swine flu vaccine allegedly caused Guillain-Barré syndrome), *aff'd in part*, 830 F.2d 831 (8th Cir. 1987); Marder v. G.D. Searle & Co., 630 F. Supp. 1087, 1092 (D. Md. 1986) (pelvic inflammatory disease allegedly caused by Copper 7 IUD), *aff'd without op. sub nom.* Wheelahan v. G.D. Searle & Co., 814 F.2d 655 (4th Cir. 1987); *In re* "Agent Orange" Prod. Liab. Litig., 597 F. Supp. 740, 835–37 (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), *cert. denied*, 484 U.S. 1004 (1988); Cook v. United States, 545 F. Supp. 306, 308 (N.D. Cal. 1982) (swine flu vaccine allegedly caused Guillain-Barré syndrome); Landrigan v. Celotex Corp., 605 A.2d 1079, 1087 (N.J. 1992) (relative risk greater than 2.0 "support[s] an inference that the exposure was the probable cause of the disease in a specific member of the exposed population"); Merrell Dow Pharms., Inc. v. Havner, 953 S.W.2d 706, 718 (Tex. 1997) ("The use of scientifically reliable epidemiological studies and the requirement of more than a doubling of the risk strikes a balance between the needs of our legal system and the limits of science."). *But cf. In re* Fibreboard Corp., 893 F.2d 706, 711–12 (5th Cir. 1990) (The court disapproved a trial in which several representative

An alternative, yet similar, means to address probabilities in individual cases is use of the attributable risk parameter.[141] The attributable risk is a measurement of the excess risk that can be attributed to an agent, above and beyond the background risk that is due to other causes.[142] When the attributable risk exceeds 50% (equivalent to a relative risk greater than 2.0), this logically might lead one to believe that the agent was more likely than not the cause of the plaintiff's disease.

The discussion above contains a number of assumptions: that the study was unbiased, sampling error and confounding were judged unlikely or minimal, the causal factors discussed in section V point toward causation, and the relative risk found in the study is a reasonably accurate measure of the extent of disease caused by the agent. It also assumes that the plaintiff in a given case is comparable to the subjects who made up the exposed cohort in the epidemiologic study and that there are no interactions with other causal agents.[143]

Evidence in a given case may challenge one or more of those assumptions. Bias in a study may suggest that the outcome found is inaccurate and should be estimated to be higher or lower than the actual result. A plaintiff may have been exposed to a dose of the agent in question that is greater or lower than that to which those in the study were exposed.[144] A plaintiff may have individual factors, such as higher age than those in the study, that make it less likely that

cases would be tried and the results extrapolated to a class of some 3,000 asbestos victims, without consideration of any evidence about the individual victims. The court remarked that under Texas law, general causation, which ignores any proof particularistic to the individual plaintiff, could not be substituted for cause in fact.).

141. *See supra* § III.C.

142. Because cohort epidemiologic studies compare the incidences (rates) of disease, measures like the relative risk and attributable risk are dependent on the time period during which disease is measured in the study groups. Exposure to the agent may either accelerate the onset of the disease in a subject who would have contracted the disease at some later time—all wrongful death cases entail acceleration of death—or be the cause of disease that otherwise would never have occurred in the subject. This creates some uncertainty (when pathological information does not permit determining which of the foregoing alternatives is the case) and ambiguity about the proper calculation of the attributable risk, that is, whether both alternatives should be included in the excess risk or just the latter. *See* Sander Greenland & James M. Robins, *Conceptual Problems in the Definition and Interpretation of Attributable Fractions*, 128 Am. J. Epidemiology 1185 (1988). If information were available, the legal issue with regard to acceleration would be the characterization of the harm and the appropriate amount of damages when a defendant's tortious conduct accelerates development of the disease. *See* Restatement (Second) of Torts § 924 cmt. e (1977); Keeton et al., *supra* note 107, § 52, at 353–54; Robert J. Peaslee, *Multiple Causation and Damages*, 47 Harv. L. Rev. 1127 (1934).

143. *See* Greenland & Robins, *supra* note 142, at 1193.

144. *See supra* § V.C; *see also* Ferebee v. Chevron Chem. Co., 736 F.2d 1529, 1536 (D.C. Cir.) ("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."), *cert. denied*, 469 U.S. 1062 (1984); *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).

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. Pathological-mechanism evidence may be available for the plaintiff that is relevant to the cause of the plaintiff's disease.[145] 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 similar) factors is required.[146]

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.[147] For example, genetics might be known to be responsible for 50% of the incidence of a disease independent of exposure to the agent.[148] If genetics can be ruled out 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.[149]

145. *See* Tobin v. Astra Pharm. Prods., Inc., 993 F.2d 528 (6th Cir.) (plaintiff's expert relied predominantly on pathogenic evidence), *cert. denied*, 510 U.S. 914 (1993).

146. *See* Merrell Dow Pharms., Inc. v. Havner, 953 S.W.2d 706, 720 (Tex. 1997); Mary Carter Andrues, Note, *Proof of Cancer Causation in Toxic Waste Litigation*, 61 S. Cal. L. Rev. 2075, 2100–04 (1988). An example of a judge sitting as fact finder 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).

147. *See, e.g.*, Grassis v. Johns-Manville Corp., 591 A.2d 671, 675 (N.J. Super. Ct. App. Div. 1991): "The physician or other qualified expert may view the epidemiological studies and factor out other known risk factors such as family history, diet, alcohol consumption, smoking . . . or other factors which might enhance the remaining risks, even though the risk in the study fell short of the 2.0 correlation." *See also In re* Joint E. & S. Dist. Asbestos Litig., 52 F.3d 1124 (2d Cir. 1995) (holding that plaintiff could provide sufficient evidence of causation without proving a relative risk greater than 2); *In re* Joint E. & S. Dist. Asbestos Litig., 964 F.2d 92, 97 (2d Cir. 1992), *rev'g* 758 F. Supp. 199, 202–03 (S.D.N.Y. 1991) (requiring relative risk in excess of 2.0 for plaintiff to meet burden of production); Jones v. Owens-Corning Fiberglas Corp., 672 A.2d 230 (N.J. Super. Ct. App. Div. 1996).

148. *See In re* Paoli R.R. Yard PCB Litig., 35 F.3d 717, 758–59 (3d Cir. 1994) (discussing the technique of differential diagnosis to rule out other known causes of a disease for a specific individual).

149. 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). To date, courts have not adopted a rule that would apportion damages based on the probability of cause in fact in toxic substances cases.

Glossary of Terms

The following terms and definitions were adapted from a variety of sources, including A Dictionary of Epidemiology (John M. Last et al. eds. 3d ed. 1995); 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 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*% of

lung cancer is attributable to cigarettes). Synonymous terms include attributable fraction, attributable risk, and etiologic fraction. See attributable risk.

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.

background risk of disease. Background risk of disease (or background rate of disease) is the 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, beta 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. Causation, as we use the term, denotes an event, condition, characteristic, or agent's 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 probability 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, follow-up 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 the effects of two processes are not separated. The distortion can lead to an erroneous result.

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, individuals in a study are incorrectly determined to be 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. 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 characteristic 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. An error that occurs when a correlation between an agent and disease in a group (ecological) is not reproduced when individuals are studied. For example, at the ecological (group) level, a correlation has been found in several studies between the quality of drinking water and mortality rates from heart disease; it would be an ecological fallacy to infer from this alone that exposure to water of a particular level of hardness necessarily influences the individual's chances of contracting or dying of heart disease.

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.

follow-up study. See cohort study.

general causation. General causation is concerned with 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. A study is generalizable when the results 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 affect the rate of disease in the populations. The rate of disease in a large reference population, such as all residents of a country, is calculated and adjusted for any differences in age between the reference population and the study population. This adjusted rate is compared with the rate of disease in the study population and provides a standardized mortality (or morbidity) ratio, which is 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. Risk factors interact, or there is interaction among risk factors, 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 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. Morbidity is the state of illness or disease. Morbidity rate may refer to the incidence rate or prevalence rate of disease.

mortality rate. Mortality refers to death. The mortality rate expresses the 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 analytic method that allows the simultaneous study of two or more independent factors or variables.

nondifferential misclassification. A form of bias that is due to misclassification of individuals or a variable of interest into the wrong category when the misclassification varies among study groups. This bias may result from limitations in data collection 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.

pathognomonic. An agent is pathognomonic when it 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.

***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.

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. In a prospective study, 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).

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 fifty 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, specificity. Sensitivity measures 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 1,000 women have cervical cancer. If the entire group of 1,000 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%. Specificity measures 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 1,000 women do not have cervical cancer. If the entire group of 1,000 women is screened for cervical cancer and the screening test only identifies 900 women as without cervical cancer, the screening test has a specificity of 900/980, or 92%.

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.

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 standard or known 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. The process of or result of separating a sample into several subsamples according to specified criteria, such as age or socioeconomic status. Researchers may control the effect of confounding variables by stratify-

ing the analysis of results. For example, lung cancer is known to be associated with smoking. To examine the possible association between urban atmospheric pollution and lung cancer, the researcher may divide the population into strata according to smoking status, thus controlling for smoking. The association between air pollution and cancer then can be appraised separately within each stratum.

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.

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 item 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., 3d ed. 1995).

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

Joseph L. Fleiss, Statistical Methods for Rates and Proportions (1981).

Leon Gordis, Epidemiology (2d ed. 2000).

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

Harold A. Kahn, An Introduction to Epidemiologic Methods (1983).

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).

Judith S. Mausner & Anita K. Bahn, Epidemiology: An Introductory Text (1974).

Marcello Pagano & Kimberlee Gauvreau, Principles of Biostatistics (1993).

Richard K. Riegelman & Robert A. Hirsch, Studying a Study and Testing a Test: How to Read the Health Science Literature (3d ed. 1996).

Bernard Rosner, Fundamentals of Biostatistics (4th ed. 1995).

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

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

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

References on Law and Epidemiology

2 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*, [1997] 27 Envtl. L. Rep. (Envtl. L. Inst.) 10279 (June1997).

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 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).

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).

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).

2 Modern Scientific Evidence: The Law and Science of Expert Testimony (David L. Faigman et al. eds.,1997).

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

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).

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).

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

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