

Reference Guide on Epidemiology

Linda A. Bailey
Leon Gordis
Michael Green

With contribution of case citations by Paul Rothstein

Linda A. Bailey, M.H.S., J.D., is Program Director, Institute for Injury Reduction, Crofton, Maryland. **Leon Gordis**, M.D., Dr.P.H., is Professor, Department of Epidemiology, Johns Hopkins School of Hygiene and Public Health, Baltimore, Maryland, and Department of Pediatrics, Johns Hopkins School of Medicine. **Michael Green**, J.D., is Professor of Law, University of Iowa College of Law, Iowa City, Iowa. **Paul Rothstein**, J.D., is Professor of Law, Georgetown University Law Center, Washington, D.C.

Contents

- I Introduction 125
- II. Trustworthiness of Research Methods 129
 - A. Was the Research Design Appropriate for Answering the Research Question? 131
 - 1. Cohort studies 134
 - 2. Case-control studies 136
 - B. Were the Study Populations Well Defined and Samples Adequately Selected So As to Allow for Meaningful Comparisons (Between Study Groups or Between Time Periods)? 138
 - 1. Did the researcher minimize the risk of selection bias? 138
 - a. How were the cases and controls (in case-control study) or exposed and unexposed subjects (in cohort study) identified and selected? 138
 - b. What percentage of those selected for the study agreed to participate? 139
 - c. What proportion of the subjects dropped out of the study before it was completed? 139
 - 2. Was the sample size adequate to draw a valid conclusion? 139
 - a. Sample size calculation 140
 - b. Power calculation 141
 - C. Was Exposure to the Putative Agent Measured Using a Standardized and Reliable Methodology? 143
 - 1. Were data collected from objective and reliable sources? 144
 - 2. What types of procedures were instituted to control the quality of measurements of exposure? 145
 - 3. Was information obtained from one group of the study population more accurate or complete than that obtained from the comparison group? 145
 - 4. Did the method of collecting data yield reliable information? 145
 - D. Were the Health Effects (i.e., Disease, Disability) Clearly Defined and Reliably Measured? 146

III.	Association Between Exposure and the Disease	147
A.	What Is the Basis for Concluding That the Exposure Is Associated with an Increased Risk of Disease?	147
1.	Relative risk (RR)	147
2.	Odds ratio (OR)	149
3.	Attributable proportion of risk (APR)	149
B.	What Categories of Error Might Have Produced a False Result?	150
C.	What Statistical Methods Exist to Evaluate the Likelihood That the Result of an Epidemiological Study Was Due to Random Sampling Error?	151
1.	False positive error and statistical significance	152
2.	False negative error	155
3.	Power	156
D.	What Biases May Have Existed That Would Result in an Erroneous Association?	156
IV.	General Causal Association Between Exposure and the Disease	157
A.	Could a Confounding Factor Be Responsible for the Study Result?	158
1.	What techniques, if any, were used to identify confounding factors?	159
2.	What techniques, if any, were used to control confounding factors?	160
B.	Overall, Does Application of the Guidelines for Causation Support a Finding of Causation?	160
1.	How strong is the association between the exposure and disease?	161
2.	Is there a temporal relationship?	162
3.	Is the association consistent with other research?	162
4.	Is the association biologically plausible (consistent with existing knowledge)?	163
5.	Have alternative explanations been ruled out?	163
6.	Does the association exhibit specificity?	163
7.	Is there a dose-response relationship?	164
C.	What Type of Causal Association Has Been Demonstrated Between Exposure and Disease?	164
V.	The Role of Epidemiology in Proving Individual Causation	167
	Glossary of Terms	171
	References on Epidemiology	179

I. Introduction

Epidemiology is the field of public health that studies the *incidence*, distribution, and etiology of disease in human populations and applies the findings to alleviate health problems. 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 are at increased *risk* of contracting particular diseases.

Judges and juries increasingly are presented with epidemiological evidence as the basis of an expert's opinion. Judges determine whether such evidence, or the expert's opinion that relies on epidemiology, reaches the jury.¹ When judges are unclear about how to gauge the quality of the expert's science, and hence the validity of the expert's testimony, incorrect and inconsistent judgments may result.²

In the courtroom epidemiological research findings are offered to establish or dispute whether *exposure* to an agent caused a harmful effect or disease.³

1. An epidemiological study, which often is published in a medical or scientific journal, is hearsay. An epidemiological 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, Fed. R. Evid. 803(18), to the hearsay rule. See *Ellis*, 745 F.2d at 305, 306 & n.18, or possibly the catchall exceptions, Fed. R. Evid. 803(24) and 804(5).

In any case, an epidemiological 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) (epidemiological study offered in evidence to support expert's opinion under New Jersey evidentiary rule equivalent to Fed. R. Evid. 703).

2. Compare *Richardson v. Richardson-Merrell, Inc.*, 857 F.2d 823 (D.C. Cir. 1988) (affirming the grant of a judgment n.o.v. to defendant in a Bendectin case), *cert. denied*, 493 U.S. 882 (1989) with *Oxendine v. Merrell Dow Pharmaceuticals, Inc.*, 506 A.2d 1100 (D.C. 1986) (reinstating a \$750,000 jury verdict for plaintiff in a Bendectin case), *cert. denied*, 493 U.S. 1074 (1990); see also *Turpin v. Merrell Dow Pharmaceuticals, Inc.*, 959 F.2d 1349, 1349 (6th Cir.) ("the inconsistent results reached by courts and juries nationwide on the question of causation in Bendectin birth defect cases are of serious concern"), *cert. denied*, 113 S. Ct. 84 (1992).

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

Epidemiological evidence identifies agents that are associated with an increased risk of disease in groups of individuals, quantifies the amount of excess disease that is associated with an agent, and provides a profile of the type of individual who is likely to contract a disease after being exposed to an agent.⁴ 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 this individual?).⁵ For example, in the 1950s Doll and Hill published a series of articles about the increased risk of lung cancer in cigarette smokers. Their findings 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.⁶ Doll and Hill's study identified an *association* between smoking cigarettes and death from lung cancer.

Association is not *causation*.⁷ An association identified in an epidemiological study may or may not be causal.⁸ Properly designed and executed studies enable epidemiologists to assess the existence (and strength) or absence of an association between an agent and a disease. Epidemiologists commonly use a measure called *relative risk* (RR) to indicate the strength of association between exposure and disease.⁹ A strong association that is demonstrated consistently in a series of research projects leads a researcher to infer that a causal relationship exists. Even the best of studies do not demonstrate more than a high probability of a causal relationship between exposure to an agent and a disease. In the absence of an understanding of the biological and pathological mechanisms by which disease develops, epidemiological evidence is the most valid type of scientific evidence of toxic causation.¹⁰

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*, 112 S. Ct. 1280 (1992); *Kehm v. Procter & Gamble Co.*, 580 F. Supp. 890, 898–902 (N.D. Iowa 1982) (toxic shock syndrome from use of Rely tampons), *aff'd sub nom. Kehm v. Procter & Gamble Mfg. Co.*, 724 F.2d 613 (8th Cir. 1983).

4. An *agent* is a factor, such as a drug, a microorganism, a chemical substance, or a form of radiation, whose presence or absence can result in the occurrence of a disease. A disease can have a single agent, a number of independent alternative agents, or a complex of two or more factors whose combined presence is necessary for the development of the disease. Agents are also referred to as risk factors of a disease. A *Dictionary of Epidemiology* 4 (John M. Last ed., 1988).

5. See *infra* § V for a discussion of specific causation.

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

7. 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. Epidemiological methods cannot prove causation; however, scientific evidence can lead an epidemiologist to infer that a certain agent causes a disease.

8. See *infra* §§ IV.A–IV.B.

9. See *infra* § III.A for a discussion of relative risk and other measures of risk.

10. *Richardson v. Richardson-Merrell, Inc.*, 857 F.2d 823, 830 (D.C. Cir. 1988) (epidemiology more probative than other forms of scientific studies), *cert. denied*, 493 U.S. 882 (1989); *Conde v. Velsicol Chem.*

An expert's opinion on causation in court is based on a series of epidemiological findings. 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.¹¹ The epidemiological studies that form the basis of the expert's testimony should examine persons who represent the general population or the subgroup that is of concern to the court and should assess the risk of disease with a study methodology and statistical measures that limit the opportunity for invalid findings.¹² While the findings of epidemiology always involve a measure of uncertainty, systematic methods for assessing the characteristics of persons included in the study and their risk of disease can be used to help rule out known sources of *bias* and *error*.

The epidemiologist uses *sample size* calculations and inclusion and exclusion criteria for identifying exposed and unexposed study groups (or cases and controls) to reduce potential error and bias in a study. These methods and techniques of epidemiology provide a means of assessing the relationship between a disease and its causes. Unfortunately, these tools are incapable of discerning every association, and the absence of an association should not be interpreted to mean causation does not exist. Even in a well-designed and well-analyzed study, lack of an association may only mean that (1) the sample size was not large enough to detect a weak association, or (2) the disease has multiple causes (epidemiological methods are best able to identify a single cause of disease).

As a final caveat about the limitations of epidemiology, the precision of epidemiological methods is based on the stability of studying large numbers of people. There should be a sufficiently large number of subjects, so that a small change in the number of people with the disease does not appreciably affect the results of the study. Applying population-based results to an individual plaintiff is generally beyond the limits of epidemiology. Measurements of error and risk, the hallmarks of epidemiology, lose their meaning when they are applied to an

Corp., 804 F. Supp. 972, 1025–26 (S.D. Ohio 1992) (“Epidemiologic studies are the primary generally accepted methodology for demonstrating a causal relation between a chemical compound and a set of symptoms or a disease.”), *aff'd*, No. 93-3092, 1994 U.S. App. LEXIS 10752 (6th Cir. May 16, 1994); DeLuca v. Merrell Dow Pharmaceuticals, Inc., 791 F. Supp. 1042, 1045 (D.N.J. 1992); *In re* “Agent Orange” Prod. Liab. Litig., 611 F. Supp. 1223, 1239 (E.D.N.Y. 1985) (epidemiology “assume[s] a role of critical importance”), *aff'd*, 818 F.2d 187 (2d Cir. 1987), *cert. denied*, 487 U.S. 1234 (1988).

11. See DeLuca v. Merrell Dow Pharmaceuticals, Inc., 911 F.2d 941, 953 (3d Cir. 1990) (pediatric pharmacologist expert's credentials sufficient pursuant to Fed. R. Evid. 702 to interpret epidemiological studies and render an opinion based thereon); Rubanick v. Witco Chem. Corp., 593 A.2d 733, 736 (N.J. 1991) (expert with degree in biochemistry and experience as a cancer researcher permitted to testify about causation and in part rely on epidemiological evidence in support of opinion); 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).

12. In *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 113 S. Ct. 2786, 2796–97 (1993), the Supreme Court addressed the standard for permitting an expert witness to testify to an opinion on a scientific matter. The ultimate issue in the Court's test is whether the methodology and reasoning that form the basis for the expert's opinion are scientifically valid. See generally Margaret A. Berger, Evidentiary Framework, in this manual.

individual. Nevertheless, a substantial body of legal precedent has developed that addresses the use of epidemiological evidence to prove causation for an individual litigant through probabilistic means.¹³

The following sets of questions address technical issues that arise in considering the admissibility of, and weight to be accorded to, epidemiological research findings. Over the past fifteen years, courts frequently have confronted the use of epidemiological 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.¹⁴

The use of epidemiology in legal disputes raises three issues for consideration:

1. Were the research methods trustworthy?
2. If so, is exposure to the agent associated with disease?
3. If the agent is associated with disease, is it a causal relationship?

There is an additional legal question that arises in most *toxic substances* cases. That issue is whether and how population-based epidemiological evidence can be used to infer specific causation. Sections II through V address these four questions. Section II examines research design and planning issues from the perspective of an epidemiologist planning a study and addresses concerns about designing a methodologically valid study. Section III looks at a completed study and explains the significance of a study's findings, statistical methods for assessing the possibility of sampling error, and methodological problems that may distort the outcome of a study.¹⁵ Section IV discusses general causation, considering whether an agent is capable of causing disease. Section V examines issues of specific causation, considering whether an agent caused an individual's disease.

13. See *infra* § V.

14. 911 F.2d 941, 954 (3d Cir. 1990); see also *Smith v. Ortho Pharmaceutical Corp.*, 770 F. Supp. 1561, 1571 (N.D. Ga. 1991) (explaining increased reliance of courts on epidemiological evidence in toxic substances litigation).

15. For a more in-depth discussion of the statistical basis of epidemiology, see David H. Kaye & David A. Freedman, Reference Guide on Statistics § II.D, 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 *Toxic. & Indus. Health* 93 (1985). See also References on Epidemiology and References on Law and Epidemiology at the end of this reference guide.

II. Trustworthiness of Research Methods

Ethical constraints limit the research methods that the epidemiologist can use.¹⁶ For example, to determine whether cigarette smoking is associated with lung cancer, the epidemiologist would like to compare two randomly selected groups, only one of which smokes cigarettes. Using true experimental methods, the epidemiologist would select a group of individuals and randomly assign half of them to cigarette smoke exposure and half to no cigarette smoke exposure to control for any differences that exist between smokers and nonsmokers. Thus, the epidemiologist could be relatively certain that any difference observed between the groups was caused by smoke exposure.

Since it is unethical to expose a group of human beings to a known harm, true experimental methods cannot be used. Instead, the epidemiologist uses observational methods. Observational methods are limited by the fact that researchers do not control the human subjects. Rather than randomly assign the study subjects to experimental groups (e.g., one group exposed to cigarette smoke and the other group not exposed), researchers identify a group of subjects who have voluntarily (or unknowingly) exposed themselves and compare the group's rate of disease with that of an unexposed group. Important factors that cannot be controlled directly by the epidemiologist include genetic background, lifestyle choices, and the amount and duration of exposure. These factors may be distributed differentially between the groups through random chance or some connection between exposure status and the other factors. The epidemiologist attempts to control and assess the influence of these factors through *research design* and statistical analysis.¹⁷

In addition to observational epidemiology, *toxicology* models based on animal studies (*in vivo*) may be used to determine toxicity in humans.¹⁸ Animal studies

16. Experimental studies with human beings are ethically proscribed where the agent is known or thought to be toxic. See *Ethyl Corp. v. United States Env'tl. 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 Pharmaceuticals*, 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).

17. True experimental studies require random assignment of subjects to groups. With the exception of controlled clinical trials (i.e., the type of studies used to test the effectiveness of new drug treatments), few epidemiological studies use true experimental methods. This reference guide focuses on observational studies.

18. For an in-depth discussion of toxicology, see Bernard D. Goldstein & Mary Sue Henifin, Reference

have a number of advantages. They can be conducted as experiments, and researchers control all aspects of the animals' lives. This avoids the problem of confounding,¹⁹ which epidemiology often confronts. Exposure can be carefully controlled and measured. Ethical limitations are diminished and animals can be sacrificed, which may improve the accuracy of disease assessment. Animal studies often provide useful information about pathological mechanisms and play a complementary role to epidemiology by assisting in framing hypotheses and in developing study designs for epidemiological studies.

Animal studies, however, have two significant disadvantages. First, animal study results must be extrapolated to another species—human beings—where 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.²⁰ 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.²¹ Those matters are almost always fraught with considerable, and currently unresolvable, uncertainty.

Toxicologists also use *in vitro* methods, in which human or animal cells or tissue 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 tissues in laboratories to whole human beings.²²

Often toxicological studies are the only or best available evidence of toxicity. Epidemiological studies are difficult, time-consuming, and expensive and consequently do not exist for a large array of environmental agents. Where both animal toxicology and epidemiological studies are available, no universal rules exist for how to interpret or reconcile them.²³ Careful assessment of the method-

Guide on Toxicology, in this manual.

19. See *infra* § IV.A.

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

21. See *infra* § IV.B.7 and accompanying note.

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

23. 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. 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).

ological *validity* and power²⁴ of the epidemiological evidence must be undertaken as well as consideration of the quality of the toxicology studies and the questions of interspecies extrapolation and dose-response relationship.

When reviewing the methodological validity of an epidemiological study, four issues should be considered:

1. Was the research design appropriate for answering the research question?
2. Were the study populations well defined and *samples* adequately selected so as to allow for meaningful comparisons (between study groups or between time periods)?
3. Was exposure to the putative agent measured using a standardized and reliable methodology?
4. Were the health effects (i.e., disease, disability) clearly defined and reliably measured?²⁵

A. Was the Research Design Appropriate for Answering the Research Question?

Research begins with formulation of the research question. This question should be stated clearly by the researcher before the data collection begins, since the researcher cannot measure the uncertainty or potential for error when the findings are unrelated to the research question. Unrelated findings may have some validity and therefore be relevant to a disputed issue in court. However, such findings should be carefully examined for bias.²⁶

In reviewing the research methods used to conduct an epidemiological study, the potential for bias should be considered. 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. This differs from con-

One explanation for these conflicting lines of cases may be that animal toxicology has much less probative value where a substantial body of epidemiological evidence that addresses the causal issue is available. That was the case, for example, in the Bendectin cases of *Richardson, Brock*, and *Cadarian*. Where epidemiological 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 Pharmaceuticals, Inc.*, 959 F.2d 1349, 1359 (6th Cir.), *cert. denied*, 113 S. Ct. 84 (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 epidemiological evidence should be required in mass exposure cases but not in isolated exposure cases). See also Bernard D. Goldstein & Mary Sue Henifin, Reference Guide on Toxicology § I.F, in this manual.

24. See *infra* §§ II.B.2.b, III.C.3.

25. These questions are based on guidelines for assessing epidemiological studies: 3 U.S. Envtl. Protection Agency, Air Quality Criteria for Particulate Matter and Sulfur Oxides § 14.1.2 (1982).

26. See *infra* § III.

ventional (and legal) usage in which bias refers to a partisan point of view.²⁷ Bias refers to anything (other than random sampling error)²⁸ that results in error in a study and thereby compromises its validity. Bias in research can result from a defect in the design or conduct of a study. Although dozens of biases have been catalogued,²⁹ the two main classes of bias are selection bias (differences in the characteristics between the individuals who are selected for study and those who are not)³⁰ and information bias (a flaw in measuring exposure or disease between study groups).

No epidemiological study is perfect; all have some degree of bias that may affect the outcome. Some studies may be so flawed as to be virtually worthless. Finding the bias, however, can be difficult if not impossible. In reviewing the validity of an epidemiological study, the epidemiologist must identify potential biases and analyze (or use educated estimates of) the amount of error that might have been induced by the existence of the bias. Moreover, the direction of error can often be determined; depending on the specific type of bias, it may exaggerate the real association, dilute it, or even completely mask it.

A type of bias that occurs in the formulation of the research question is conceptual bias. Conceptual bias usually means that the research question and hypothesis are biologically implausible as a result of faulty logic, faulty premises, or mistaken beliefs on the part of the researcher. 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.³¹ Failure to have such a basis raises concerns about conceptual bias.³²

Once the research question has been identified, the researcher designs a study that elicits information directly relevant to the question. There are two classes of epidemiological research designs for studying human populations: (1) studies that collect data about the group as a whole and (2) studies that collect data about individuals within the group.

Studies that collect data about the group as a whole are called ecological studies.³³ Such studies are useful for identifying associations but generally are re-

27. A Dictionary of Epidemiology, *supra* note 4, at 14; Edmond A. Murphy, *The Logic of Medicine* 239 – 62 (1976).

28. See *infra* § III.B.

29. See David L. Sackett, *Bias in Analytic Research*, 32 *J. Chron. Dis.* 51 (1979).

30. See *infra* note 46.

31. 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.

32. The effect of this conceptual bias would be to dilute or mask any real effect that the agent might have on a specific type of birth defect. See Kenneth J. Rothman, *Modern Epidemiology* 88 (1986) (“[u]nwarranted assurances of a lack of effect can easily emerge from studies in which a wide range of etiologically unrelated outcomes are grouped.”).

33. In *Renaud v. Martin Marietta Corp.*, 749 F. Supp. 1545, 1551 (D. Colo. 1990), *aff’d*, 972 F.2d 304 (10th Cir. 1992), plaintiffs attempted to rely on an excess incidence of cancers in their neighborhood to prove

garded by epidemiologists as “weak.” An example of an ecological study follows.

If the researcher is interested in determining whether a high dietary fat intake is associated with breast cancer, he or she can compare different countries on the basis 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 findings 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). In addition to the lack of information about an individual’s intake of fat, the researcher does not know about alternative individual exposures to other agents (or family history) that may also be responsible for the increased risk of breast cancer. This lack of particularized information about an 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, known as an *ecological fallacy*. However, 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.

Another type of group study compares disease rates over time. *Secular trend studies* (also called time-line studies) focus on disease rates before and after a point in time when some event of interest took place.³⁴ 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, time-line studies are less powerful than the studies described below. Other *variables* associated with the disease, such as improved diagnostic techniques and changes in lifestyle or age demographics, may change over time. If those variables 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 studies.

Observational studies, which are research designs that collect data about individuals within a group, allow the researcher to draw stronger inferences about associations between risk factors and disease. For example, in an observational study conducted in the population described above (with a high average fat intake and an increased rate of breast cancer), the researcher gathers information

causation. Unfortunately, the court confused the role of epidemiology in proving causation with the issue of plaintiff’s exposure to the alleged carcinogen and never addressed the evidentiary value of plaintiff’s evidence of a disease cluster (i.e., the aggregation of a particular disease in a neighborhood or community). *Id.* at 1554.

34. In *Wilson v. Merrell Dow Pharmaceuticals, Inc.*, 893 F.2d 1149, 1152–53 (10th Cir. 1990), 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 defendant’s expert witnesses could rely on it in rendering their opinions.

about how much dietary fat each individual consumes and whether she has breast cancer. The researcher then compares the dietary fat intake of individuals who have breast cancer with that of those who do not to determine if fat intake is associated with breast cancer.

There are two main types of observational studies: *cohort studies* and *case-control studies*.³⁵ The difference between these two is the use of exposure or disease as the independent variable. Cohort studies, which use exposure as the independent variable, compare two groups: one group that is exposed to the agent, and a *control group* that consists of persons with similar characteristics who have not been exposed. Case-control studies compare a *case group*, those who have the disease or outcome being studied, and a control group, those who do not have the disease in question. The researcher compares the odds of having the disease when exposed to suspected agents and when not exposed. Case-control studies use disease as the independent variable.

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.

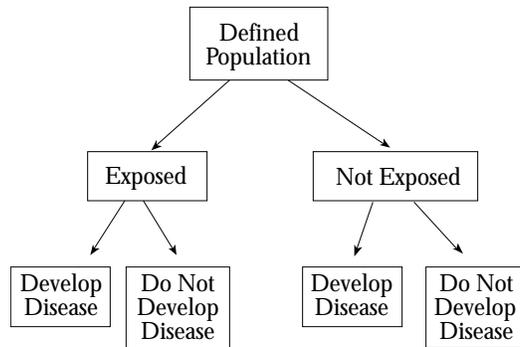
1. Cohort studies

In cohort studies (also called prospective studies, concurrent studies, follow-up studies, incidence studies, or longitudinal studies), the researcher identifies two groups of individuals: (1) individuals who have been exposed to a substance that is thought might cause a disease and (2) individuals who have not been exposed. Both groups are followed for a specified length of time, and the proportion of each group that develops the disease is compared.³⁶ If the exposure is associated with or causes the disease, the researcher would expect a greater proportion of the exposed individuals to develop the disease (see Figure 1).

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

36. Sometimes retrospective cohort studies (also known as historical cohort or retrospective follow-up 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 his 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).

Figure 1
Design of a Cohort Study



An advantage of the cohort study design is that the temporal relationship between exposure and disease can be established. By tracking the exposed and unexposed groups over time, the researcher can determine the time of disease onset. This temporal relationship is relevant 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 comprised white nonminers from the same geographic area. Members of both groups were examined every three years, and the degree of exposure of the exposed cohort to radon was measured from samples taken in the mines. The ongoing testing of rock samples for radioactivity and the periodic medical monitoring of lungs permitted the researchers to examine whether disease was linked to prior work exposure to radiation and to discern the relationship between exposure to radiation and disease. Exposure to radiation was associated with the development of lung cancer in uranium miners.³⁷

The cohort design is often used in occupational studies. A weakness of this design is that an increased risk of disease among the exposed group may be caused by agents other than the exposure. 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 it may be because low-wage groups are exposed to other harmful agents, such as environmental toxins present in higher concentrations in their neighborhood. The researcher must attempt in the study design to identify factors other than the exposure that may be responsible for the increased risk of disease. If data are gathered

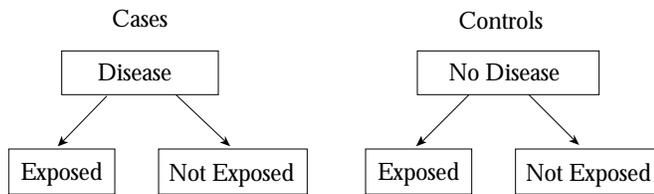
37. This example is based on a study description in Abraham M. Lilienfeld & David E. Lilienfeld, *Foundations of Epidemiology* 237–39 (2d ed. 1980). The original study is Joseph K. Wagoner et al., *Radiation as the Cause of Lung Cancer Among Uranium Miners*, 273 *New Eng. J. Med.* 181 (1965).

on other possible *etiologic factors*, the researcher can sometimes use statistical methods³⁸ 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 sections IV–IV.B.

2. Case-control studies

In case-control studies, the researcher begins with a group of individuals who have the disease (cases) and then selects a group of individuals who do not have the disease (controls). Instead of observing each group, as is done in a cohort study, the researcher compares past exposures. If a past exposure is associated with or caused the disease, the researcher expects to find a higher proportion of past exposure among the cases. For example, we expect a higher proportion of past cigarette smoking among lung cancer cases than among controls who do not have lung cancer (see Figure 2).

Figure 2
Design of a Case-Control Study



An advantage of the case-control study is that the study can be completed in less time and with less expense than a cohort study. Case-control studies also are often more powerful and therefore reveal weaker associations than cohort studies, especially when the disease or outcome is rare.³⁹

The case-control research design poses a number of potential methodological problems. However, the researcher can prevent or diminish these problems with careful attention to the design and conduct of the study. For instance, 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 subject about past exposures, thus relying on his or her memory. Research has shown that individuals with disease (cases) may more readily recall past exposures than individuals with no disease (controls);⁴⁰ this

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

39. 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).

40. Steven S. Coughlin, *Recall Bias in Epidemiologic Studies*, 43 *J. Clin. Epidemiol.* 87 (1990); Rothman, *supra* note 32, at 85.

creates a potential for biased data.

For example, consider a case-control study conducted to examine the cause of congenital malformations. The epidemiologist is interested in whether the malformation was caused by an infection during the mother's pregnancy. 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 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 differential recall between the two groups. The problem of recall bias can sometimes be overcome by finding a second source of data to validate the mother's response (e.g., blood test results from prenatal visits or medical records that document symptoms of infection).⁴¹ A cohort study would not be feasible, because malformations occur so rarely, and cohort studies may not be powerful enough to detect outcomes that are rare.

Selecting members of the control group (those without disease) also may be problematic in case-control studies, especially if these individuals differ in many of their characteristics from members of the case group (those with disease).⁴² The selection of an appropriate control group has been described as the Achilles' heel of a case-control study.⁴³ One key to a valid control group is to ensure that the controls were selected independently of their exposure status, as illustrated below.⁴⁴

Since many researchers are located in medical centers, they often select hospital patients as study participants. However, the selection of controls from a hospital's inpatient population can introduce selection bias into a study. For example, suppose an association is found between coffee drinking and coronary heart disease using hospital patients as a control group. However, the hospitalized control group may include individuals who had been advised against drink-

41. 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).

42. The types of characteristics most commonly considered when selecting the study population include age, race, socioeconomic status, years of education, and occupation.

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

44. Another important criterion for selecting controls occurs where the cases are a sample of a given population rather than all cases within the population. In that situation, care must be taken to select controls who, if they had developed the disease, would have been included as cases in the study. Thus, if the cases consist of the leukemia patients at a hematology-oncology clinic, it is important that the control group be limited to persons who, if they had contracted leukemia, would be patients (and therefore cases) at the same clinic. For additional explanation on selecting controls in case-control studies, see Brian MacMahon & Thomas F. Pugh, *Epidemiology: Principles and Methods* 244-56 (1970); Rothman, *supra* note 32, at 62-68.

ing coffee for medical reasons, such as a peptic ulcer (the reason for lower consumption is not important, but the unrepresentativeness of the control group is). If this is true, the amount of coffee drinking in the control group would understate the extent of coffee drinking in the general population. Understating the exposure to coffee in the population without disease would result in inflating the impact of exposure to coffee on heart disease. This bias must be considered when the study's findings and implications are being made. Extrapolation of the findings to the general population may not be possible (or should be done cautiously), since the control group differs in an important way (exposure to the agent).⁴⁵

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 identify whether the error tended to exaggerate or understate the true association. Thus, bias may exist in a study that nevertheless has probative value.

B. Were the Study Populations Well Defined and Samples Adequately Selected So As to Allow for Meaningful Comparisons (Between Study Groups or Between Time Periods)?

As stated above, the two main types of bias are selection bias, in which there is a systematic difference between those individuals included in the study and those who are not, and information bias, which involves error in measuring disease or exposure among those included in the study.

1. Did the researcher minimize the risk of selection bias?⁴⁶
 - a. How were the cases and controls (in case-control study) or exposed and unexposed subjects (in cohort study) identified and selected?

A list of criteria for inclusion in and exclusion from the study must be articulated by the researcher. These criteria should be documented clearly before the subjects are recruited for the study to ensure that no overt or covert biases enter into the selection process. Such biases could lead to erroneous inferences regarding causation. For example, in a *prospective study* of cervical cancer, those who are

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

46. 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. Examples include hospital cases or cases under a physician's care, excluding those who die before admission to hospital because the course of their disease is so acute, those not sick enough to require hospital care, or those excluded by distance, cost, or other factors." A Dictionary of Epidemiology, *supra* note 4, at 15.

In *In re "Agent Orange" Prod. Liab. Litig.*, 597 F. Supp. 740, 783 (E.D.N.Y. 1985), *aff'd*, 818 F.2d 145 (2d Cir. 1987), the court expressed concern about selection bias. The exposed cohort consisted of young, healthy males who served in Vietnam. Comparing mortality rates between that exposed cohort and a control group made up of civilians might have resulted in error due to selection bias. Failing to account for health status as an independent variable would tend to understate any association between exposure and disease where the exposed cohort is healthier.

not at risk for the disease—women who have had their cervixes removed and men—should be excluded from the study population. Inclusion of such individuals as controls in a case-control study could result in erroneous findings by overstating the association between the agent and the disease. If the study population is not described precisely by the researcher, the ability to generalize the results is compromised.

b. What percentage of those selected for the study agreed to participate?

Even when a study is well designed, bias may be introduced if a large proportion of those selected as subjects refuse to participate. Many studies have shown that individuals who participate in studies differ significantly from those who do not. Consequently, if only a small proportion of selected subjects agree to participate, the findings may not apply to the general population.

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 interviewing those who refused to participate or by comparing the relevant characteristics of those who refused to participate with those who participated.

c. What proportion of the subjects dropped out of the study before it was completed?

Many of the same issues discussed above apply here. If, for example, a significant number of subjects drop out of a study before completion, it may be impossible to extrapolate the findings from a small number of subjects to the general population. The researcher should examine whether the study group is still representative of the general population.

2. Was the sample size adequate to draw a valid conclusion?

Common sense leads one to believe that researchers who do not study a large enough sample of individuals may not be able to discern the relationship between exposure to a substance and a disease. Common sense also leads one to believe that by enlarging the sample size (size of the study group), researchers can form a more accurate conclusion and reduce the chance of error in their results. Both statements are correct: researchers can increase the accuracy of the measurement of the risk of disease by enlarging the sample size. This common-sense intuition is illustrated by a test to determine if a two-sided coin is fair. A test in which the coin is flipped 500 times is much more helpful than a test in which the coin is flipped 10 times. Both common sense and statistics reveal that it is far more likely that 80% of the flips in the latter test will result in heads than in the former test if the coin is fair. The concern with the design of the coin test, as with epidemiology, is that both involve sampling techniques to draw an infer-

ence. Estimates based on samples are subject to random error, which can be reduced by increasing the size of the samples.

Sample size calculations are important in two circumstances. The first circumstance occurs during the planning of a study, when the researcher estimates the size and expense of the study. At this point, the researcher must determine the number of subjects that will have to participate to obtain research findings of acceptable precision. Since enlarging the study size increases the time, cost, and complexity of conducting the study, a balance must be maintained between the scientific precision of the findings and the cost of the project.

The second circumstance occurs after a study has been completed. The outcome of a study can be incorrect or inaccurate because of sampling error. Thus, a study erroneously can find no association between exposure to an agent and a disease if the sample size is too small to detect the association that existed.⁴⁷ If the researchers suspect that this is the case after a study has been completed, they can determine the likelihood that the size of the sample (i.e., the number of participants) was sufficient to permit detection of an association of a given magnitude, if there actually was one. Similarly, a study can find an association that is spurious—the result of random error. This is similar to the example mentioned above, in which a fair coin flipped ten times results in eight heads. Statistical techniques can be used to estimate the likelihood that the association is due to sampling error.⁴⁸

Researchers use a variety of approaches to determine an appropriate size for a study population, including a sample size calculation and a power calculation. Although these calculations generally are completed before the study begins, they require an estimation of the study's findings. In general, the calculations help determine whether a study is feasible (i.e., whether the researcher can recruit enough subjects and finance the project adequately). It should be recognized that sample size calculations are based on public health considerations and costs, which may not coincide with the level of precision that would be optimal for legal standards of proof.

a. Sample size calculation

The sample size calculation provides researchers with an estimate of the number of individuals they should study to detect whether exposure to a substance increases the risk of disease. The calculation is based on four factors:

1. the specified level of *statistical significance*, or *alpha*, that is desired;⁴⁹

47. Junius C. McElveen, Jr. & Pamela S. Eddy, *Cancer and Toxic Substances: The Problem of Causation and the Use of Epidemiology*, 33 Clev. St. L. Rev. 29, 40–41 (1984) (detecting an increase of 200 cancers per population of 100,000, with a significance level of .05, where the background rate of cancer is 20,000 per 100,000 individuals, would require exposed and control cohorts of 700,000 persons each).

48. See *infra* § III.C.1.

49. The specified level of statistical significance, also called alpha or type I error, is the probability of observing an association of the magnitude found in the study or greater when there is no association (i.e., false positive). See *infra* § III.C.1.

2. the chance of missing a real effect, or beta, that the researcher selects;⁵⁰
3. the estimated magnitude of the increased risk of disease, or *effect size* ;⁵¹
and
4. the *background risk of disease* or exposure.⁵²

Factors (1) and (2) are set by convention and generally do not change from study to study. Changes in these values should be based on logical and defensible scientific needs. For example, it may be important to increase beta or alpha to examine carefully the effect of a specific risk factor.

Factors (3) and (4) are critical in determining an adequate sample size. For example, in a study of exposure to video display terminals (VDTs) and spontaneous abortion, the researcher must consider two questions. First, what is the background rate of spontaneous abortion in the group? Second, how many excess spontaneous abortions are thought to be related to VDT exposure? In general, when there is a high background rate of a particular disease and when the increased risk of disease is small, the researcher needs a larger sample size. In the example given above, the researcher would need a fairly large sample size, since the background rate of spontaneous abortions is high and other studies suggest that the risk associated with VDT exposure is small.

Since the sample size calculation generally is performed before the study is initiated, the researcher often estimates values for the increased risk of disease and the standard deviation.⁵³ Researchers can rely on values from similar studies conducted, or if no such studies are available, they can rely on educated guesswork. When little is known about a particular disease, researchers can calculate the sample size and then increase it to allow for the uncertainty.

This method of calculating sample size has been criticized for being subject to manipulation, because the researcher must estimate values that will not be known until the study has been completed. Nonetheless, it still is used for determining the size of a study population.

b. Power calculation

The results of a power calculation, often displayed as a diagram, present the probability that a researcher will be able to find a hypothetical increased risk of disease for specified sample sizes. After reviewing the diagram, a researcher

50. The chance of missing a real effect, or beta, is the probability that an association that exists will be missed by the study. See *infra* § III.C.2.

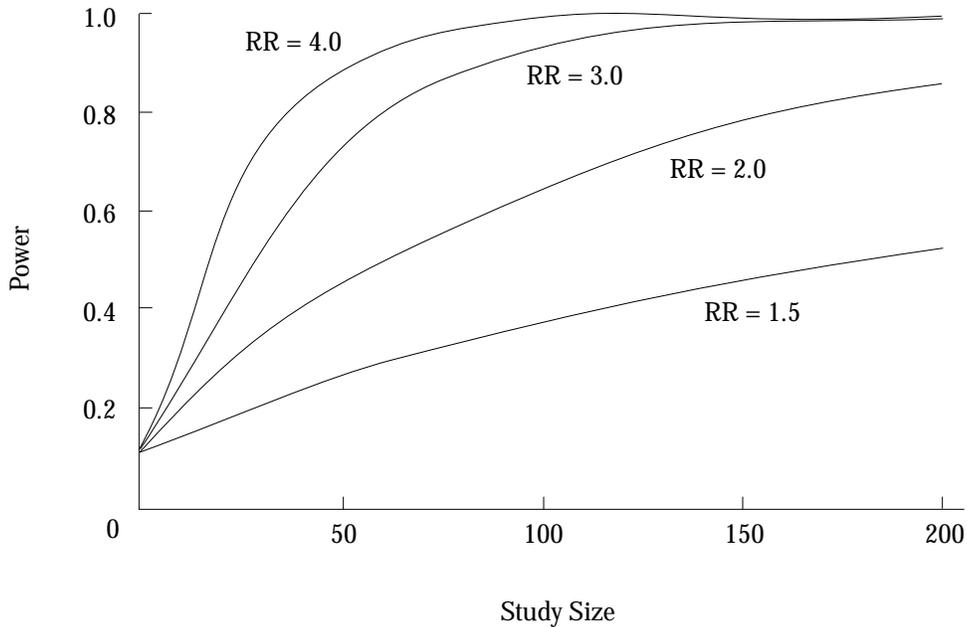
51. The magnitude of the increased risk in disease, or effect size, is best thought of as the amount of disease that is caused by exposure to a toxic substance. For example, the risk of contracting lung cancer may be ten times higher for cigarette smokers than for nonsmokers. The magnitude of the increased risk is therefore a factor of 10. See *infra* § III.A.

52. The background rate, or background risk, in a population is the amount of disease that occurs in individuals who have no known exposures to an alleged risk factor for the disease. For example, the background rate for all birth defects is 3%–5% of live births.

53. Standard deviation is a summary statistic that describes how widely dispersed the data are around the mean (average) value.

chooses the appropriate sample size. Figure 3 shows sample power curves, with *power* plotted against sample size for several anticipated levels of relative risk (RR, a measure of association discussed below). Many statistical computer programs used in epidemiology can generate power curves.

Figure 3
Power Curves for a Case-Control Study



The power curves in Figure 3 are drawn on the assumption that the researcher will conduct a case-control study that will have an equal number of cases (those with disease) and controls (those without disease). The estimated exposure rate in the control group is 0.3, which means that 30% of the controls are predicted to have been exposed to the agent. In addition, the curves are calculated and drawn based on a level of statistical significance of 0.1. The y-axis displays the power, or probability (with 1.0 being 100%), of the study being able to detect an association of the magnitude shown for each of the curves based on the sample size on the x-axis. Thus, if the researcher wants to detect a relative risk of at least 3.0 with 80% probability, there should be approximately 75 subjects in each of the case and control groups.

After a study is completed, a power curve is used to determine the likelihood that the study would have detected an association of a given magnitude. Thus, a case-control study with 100 subjects in each group has a slightly better than 60% probability (determined from reading the curve) of detecting a relative risk of 2.0

that will be statistically significant. Put another way, there is a 40% chance that the study failed to detect a relative risk of up to 2.0 at a statistical significance level of 0.1.

Power calculations performed in advance of a study have also been criticized as being subject to manipulation. They can be manipulated because the researcher must estimate study variables in advance of the study. Nevertheless, power calculations are becoming a more common method for determining study size, in part because a power diagram provides more information than a sample size calculation.⁵⁴ This criticism is not applicable to power calculations performed after a study is completed, because at that time, exposure rates (for case-control studies) and disease incidence (for cohort studies) are known.

Both the power calculation and the sample size calculation require some estimation or educated guesswork by the researcher and are subject to uncertainty. Although there is no absolute right number of participants, the researcher usually uses one of these two approaches to determine the minimum number of participants needed. The assumptions that underlie a researcher's estimations can be examined for soundness. The researcher should be able to articulate a reasonable and scientific basis for estimates of the magnitude of the increased risk of disease and the background risk. If the study size is too small, or if the actual risk from exposure is less than the researcher estimated, the study may be inconclusive.

C. Was Exposure to the Putative Agent Measured Using a Standardized and Reliable Methodology?

One of the most difficult areas in epidemiology concerns exposure: determining whether a person was exposed to an agent in the past and, if so, measuring the intensity and length of such an exposure.⁵⁵ Exposure can be measured directly or indirectly.⁵⁶ Sometimes researchers use a *biological marker* as a direct measure of exposure—an alteration in tissue or body fluids that occurs as a result of

54. Rothman, *supra* note 32, at 79–81.

55. Dose generally refers to the intensity or magnitude of exposure multiplied by the time exposed. For a discussion of the difficulties of determining dosage 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).

A different, but related, problem often arises in court. Determining plaintiff's exposure to the alleged toxic substance always involves a retrospective determination and may involve difficulties similar to those faced by an epidemiologist planning a study. Thus, in *Christophersen v. Allied-Signal Corp.*, 939 F.2d 1106, 1113 (5th Cir. 1991), *cert. denied*, 112 S. Ct. 1280 (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 plaintiff demonstrate (1) regular use by an employer of defendant's asbestos-containing product; (2) plaintiff's proximity to those products; 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).

56. 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).

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 or not a person was exposed. Biological markers rarely help determine the intensity or duration of exposure.⁵⁷

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 installing 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. Where the agent of interest is a drug, medical or hospital records can be used to determine exposure. Thus, retrospective occupational or environmental measurements of exposure are usually less accurate than prospective or follow-up studies, especially ones where drugs 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.⁵⁸

1. Were data 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.⁵⁹ 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.

57. The timing of exposure may also be critical, especially where the disease of interest is birth defects. 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: This definition of exposure is too broad because environmental agents are only likely to cause birth defects during a narrow band of time.

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

59. 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).

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.

2. What types of procedures were instituted to control the quality of measurements of exposure?

The types of quality control procedures used depend on the source of 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 reliability of the laboratory test.

3. Was information obtained from one group of the study population more accurate or complete than that obtained from the comparison group?

Error can be introduced into a study if there are differences in the accuracy or completeness of the subjects' recollection of past events or experiences. This type of bias, known as recall bias, is a special concern in case-control studies.

For example, a researcher may be interested in whether fetal malformation is caused by a mother's exposure to a virus during pregnancy. 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 the malformed infants may tend to recall inconsequential fevers or runny noses during pregnancy that readily would be forgotten by a mother who had a normal infant. Even if the true viral infection rate in mothers of malformed infants is no different from the rate in mothers of normal infants, the results of this study would indicate a false association between infection during pregnancy and birth defects because of differential recall between the two groups.⁶⁰

4. Did the method of collecting data yield reliable information?

Errors in data collection can compromise the validity of the research findings. Evidence of staff training and data collection guidelines may be available for review by opposing experts. If the data were coded by members of the research team, reliability checks and coefficients may be reported on the coding and data

60. See *Brock v. Merrell Dow Pharmaceuticals, 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). It should be noted 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 23, 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); *infra* § III.B.

entry processes. Further, if the data were collected through interview protocol, survey form, or code sheet, a pilot test may have been conducted using the data collection instrument. The results of the pilot test would indicate whether the data collection instruments (questionnaires, forms, etc.) posed problems for data collection staff.

D. Were the Health Effects (i.e., Disease, Disability) Clearly Defined and Reliably Measured?

The outcome or health effects being studied should be clearly defined by the researcher in the study design. Precise definition of the disease ensures that the same variable is consistently measured throughout the study. For example, if a researcher is studying birth defects, it is necessary to define the age at which defects will be measured, as some birth defects are not apparent at birth and only are diagnosed later in childhood.

The quality and sophistication of the diagnostic methods used to detect a 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.⁶¹ 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 positives (indicating pregnancy when the woman is not pregnant), the study will overestimate the number of spontaneous abortions.

61. In *In re Swine Flu Immunization Prod. Liab. Litig.*, 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.

III. Association Between Exposure and the Disease

Exposure to an agent and disease are said to be associated when they occur more frequently together than one would expect by chance.⁶² The term association implies a range of possible relationships, but it does not necessarily imply a cause-effect relationship between exposure and disease. Of course, a causal relationship is one possible explanation for the association, which is of ultimate concern to epidemiologists.

This section begins with a description of the epidemiological methods for expressing the strength of an association between exposure and disease. It goes on to review ways in which an incorrect result can be produced and then examines statistical methods for evaluating whether an association is real or due to sampling error.

A. What Is the Basis for Concluding That the Exposure Is Associated with an Increased Risk of Disease?

The strength of an association between exposure and disease can be stated as a relative risk (RR), *odds ratio* (OR), or *attributable proportion of risk* (APR). Each of these measurements of association examines the degree to which the risk of disease increases when individuals are exposed to an agent.

1. Relative risk (RR)

A commonly used approach for expressing the association between an agent and disease is relative risk. It is defined as the ratio of the incidence of disease in exposed individuals compared to the incidence in unexposed individuals.⁶³ Thus, it can be expressed algebraically as:

$$RR = \frac{I_e}{I_c}$$

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

63. This definition of relative risk assumes that the researcher is conducting a cohort study (examining the risk of disease in an exposed and an unexposed population). In a case-control study, the equivalent of the relative risk, the odds ratio, compares the odds of having disease when exposed to a suspected agent and when not exposed.

In the formula above, RR is the relative risk, I_e is the incidence of disease in the exposed population, and I_c is the incidence of disease in the control population.

For example, a researcher studies 100 individuals who are exposed to an agent and 100 who are not exposed. After several years, 40 of the exposed individuals are diagnosed as having a disease, and 10 of the unexposed individuals also are diagnosed as having disease. The relative risk of contracting the disease is calculated as follows:

- The incidence of disease in the exposed individuals is 40 cases per 100 persons (40/100), or 0.4.
- The incidence of disease in the unexposed individuals is 10 cases per 100 persons (10/100), or 0.1.
- The relative risk is calculated as the incidence in the exposed group (0.4) divided by the incidence 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 higher than the risk of disease in the unexposed group.⁶⁴

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 the agent and the disease.
- 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.

Researchers should scrutinize their results for error. Error in the design of a

64. The court in *Gaul v. United States*, 582 F. Supp. 1122, 1125 n.9 (D. Del. 1984), defined relative risk as follows:

Relative risk, or relative risk ratio, describes the relationship between the risk of an occurrence, such as contracting a disease, in a population exposed to a certain stimulus, and the risk of the occurrence in a population not exposed to the stimulus. It is the ratio of the former risk to the latter. It is another way of explaining how much more likely a person exposed to the stimulus is to get a disease than an unexposed person. For example, using hypothetical numbers and facts, if one in every 100,000 vegetarians contracts stomach cancer while five in every 100,000 meat eaters contract this disease, the relative risk of contracting cancer among meat eaters would be 5/1, or 5. In other words, the risk of getting stomach cancer would be five times greater for meat eaters than vegetarians, assuming all other factors are held constant.

See also *DeLuca v. Merrell Dow Pharmaceuticals, Inc.*, 911 F.2d 941, 947 (3d Cir. 1990) (“[i]n the context of an epidemiological study of Bendectin’s relationship to birth defects, the relative risk is the ratio of the incidence rate of birth defects in the study group exposed to Bendectin divided by the rate in the control group not exposed to Bendectin.”).

study could yield an incorrect relative risk. Sources of bias should be examined. Whenever a positive association is uncovered, further analysis should be conducted to determine if the association is causal.⁶⁵

2. Odds ratio (OR)

The odds ratio is similar to a risk ratio. An odds ratio is used for case-control studies and is based on a comparison of the odds of having a disease when exposed to a suspected agent and when not exposed. For all practical purposes, the odds ratio is comparable to the relative risk when the disease is rare.⁶⁶ However, as the disease becomes more common, these measures diverge.

The odds ratio is expressed algebraically as:

$$OR = \frac{D_e \times C_u}{D_u \times C_e}$$

In the formula above, OR is the odds ratio, D_e is the number of cases (those with the disease) who were exposed to the agent, C_u is the number of controls (those without the disease) who were not exposed to the agent, D_u is the number of cases who were not exposed to the agent, and C_e is the number of controls who were exposed.

Consider the following hypothetical study: A researcher finds 10 individuals with a disease. Four of those individuals were exposed to the agent and 6 were not. The control group consists of 100 persons, none of whom have the disease “by definition.” Among the control group, 20 have been exposed and 80 have not. The calculation of the odds ratio would be:

$$OR = \frac{4 \times 80}{6 \times 20} = 2.67$$

If the disease is relatively rare in the general population (about 5% or less), the odds ratio is close to a relative risk of 2.67, which means that there is almost a tripling of the disease in those exposed to the agent.

3. Attributable proportion of risk (APR)

Perhaps the most useful measurement of risk, the attributable proportion of risk (also called etiologic fraction and attributable risk percent) represents the proportion of the disease among exposed individuals that is associated with the exposure. The attributable proportion reflects the maximal amount of the disease that could be prevented by blocking the effect of the exposure or by eliminating the exposure.⁶⁷ In other words, if the association is causal, the attributable proportion of risk is the amount of disease in an exposed population caused by

65. See *infra* §§ IV–IV.B.

66. For further detail about the odds ratio and its calculation, see Kahn & Sempos, *supra* note 39, at 47–56.

67. Rothman, *supra* note 32, at 38–39. See also *Landrigan v. Celotex Corp.*, 605 A.2d 1079, 1086 (N.J. 1992) (illustrating that relative risk of 1.55 conforms to attributable risk of 35%).

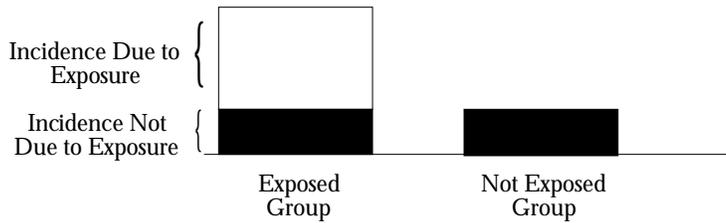
the agent (see Figure 4).

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 nonexposed group. With that information, the attributable proportion of risk can be stated algebraically as:

$$APR = \frac{l_e - l_c}{l_e}$$

In the above formula, *APR* is the attributable proportion of risk, l_e is the incidence of disease in the exposed group, and l_c is the incidence of disease in the control group.

Figure 4
Risks in Exposed and Not Exposed Groups



The attributable proportion of risk can be calculated using the example described in section III.A.1. Suppose a researcher studies 100 individuals who are exposed to a substance and 100 who are not exposed. After several years, 40 of the exposed individuals are diagnosed as having a disease, and 10 of the unexposed individuals are also diagnosed as having a disease.

- The incidence of disease in the exposed group is 40 persons in 100.
- The incidence of disease in the unexposed group is 10 persons in 100.
- The maximum proportion of disease that is attributable to the exposure is 30 persons out of 40, or 75%.

This means that up to 75% of the disease in the exposed group is attributable to the exposure.

B. What Categories 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 erroneously may conclude that there is no association. Finally, a study may find an association when one truly exists, but the association found may be greater or less than the real association.

Two categories of error in an epidemiological study can produce these incor-

rect results.⁶⁸ The first, known as sampling error, occurs because all epidemiological studies are based on sampling a small proportion of the relevant population. As stated in section II.B.2, the size of the sample can be adjusted to reduce (but not eliminate) the likelihood of sampling error. Statistical techniques⁶⁹ permit an assessment of the plausibility that the results of a study represent a true association or random error.

Systematic error or bias also can produce error in the outcome of a study. Many of the potential sources of bias were described in section II in connection with the planning or conduct of an epidemiological study. However, even the best designed and conducted studies still can have biases. Thus, after a study is completed (and this is the time when most lawyers and judges confront an epidemiological study), 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 enable an assessment of whether the study's results should be adjusted, and if so, the direction of such an adjustment and the range of error that is indicated. Sometimes, epidemiologists conduct reanalyses of a study's underlying data to correct for a bias identified in a completed study.⁷⁰

C. What Statistical Methods Exist to Evaluate the Likelihood That the Result of an Epidemiological Study Was Due to Random Sampling Error?⁷¹

Before detailing the statistical methods used to assess random error, two concepts are explained that are central to epidemiology and statistical analysis. Understanding these concepts should facilitate comprehension of the statistical methods.

Epidemiologists often refer to the *true association* (also called real association), which is the association that really exists between agent and exposure and that might be found by a perfect (but nonetheless nonexistent) study. The true association is a concept that is used in evaluating the results of a given study.

Epidemiologists begin each study with a hypothesis that they seek to dis-

68. 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.

69. See *infra* § III.C.

70. E.g., Richard A. Kronmal et al., *The Intrauterine Device and Pelvic Inflammatory Disease: The Women's Health Study Reanalyzed*, 44 J. Clin. Epidemiol. 109 (1991) (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).

71. For a bibliography on the role of statistical significance in legal proceedings, see Sanders, *supra* note 15, at 329 n.138.

prove.⁷² The hypothesis most often used is called the *null hypothesis*, which posits that there is no true association between agent and exposure; thus, the epidemiologist begins by assuming that the relative risk is 1.0 and seeks to develop data that disprove the hypothesis.⁷³

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 whether it is the result of random error. Random error is similar to the error that occurs when a fair coin yields five heads out of five tosses.⁷⁴ Thus, even though the true association is a relative risk of 1.0, an epidemiological study may find a positive association because of random error. An erroneous conclusion that the null hypothesis is false (due to random error) is a *false positive* (also, *alpha error* or *type I error*).

The essential concern is with the numerical stability of the sampling conducted by the epidemiologist. A researcher who compares two coins and finds a 50% incidence of heads in one coin and a 75% incidence of heads in the second might conclude that the second coin is biased and the first is fair. However, if each test consists of only four flips, the results are highly unstable, because if the next flip for each coin results in a tail, each one will have resulted in a 60% incidence of a head or a tail. Nothing, then, could be said about which coin is a biased one. If the test is conducted with larger numbers (1,000 flips each), the stability of the outcome is less likely to be influenced by random error, and the researcher would have greater confidence in the inferences drawn from data that found 75% heads in one coin and 50% in the other.⁷⁵

One means for evaluating the possibility that an effect is due to random error is by calculating a *p-value*.⁷⁶ A *p-value* represents the probability that a positive association like that found would result due to random error if no association is in fact present.⁷⁷ Thus, a *p-value* of .1 means that there is a 10% chance that an effect at least as large as that found is due solely to random error.⁷⁸

72. See, e.g., *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 113 S. Ct. 2786, 2796 (1993) (scientific methodology involves generating and testing hypotheses).

73. See *DeLuca v. Merrell Dow Pharmaceuticals, Inc.*, 911 F.2d 941, 945 (3d Cir. 1990).

74. See *id.* at 946-47.

75. 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 Pharmaceuticals, Inc.*, 113 S. Ct. 2786 (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 while "[s]mall communities or groups of people are deemed 'statistically unstable'" and "data from small populations must be handled with care does not mean that it cannot provide substantial evidence in aid of our effort to describe and understand events."

76. 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).

77. Technically, a *p-value* represents the probability that the study's association or a larger one would occur due to sampling error where no association (or, equivalently, the null hypothesis) is the true situation.

78. Technically, a *p-value* of .1 means that 10% of all similar studies would be expected to yield the same

To avoid false positive error, epidemiologists use a convention that the p -value must fall below some selected level known as alpha for the results of the study to be considered statistically significant.⁷⁹ The most common level of statistical significance or alpha used is .05. Using this value for significance testing accepts as statistically significant those studies where a positive association erroneously is found (because the true situation is that there is no association) no more than 5 times out of 100 due to random error.⁸⁰ Although .05 is often the level of alpha selected, other defensible levels can and have been used legitimately. Thus, in its study of the effects of secondhand smoke, the Environmental Protection Agency (EPA) used a .10 standard for statistical significance.⁸¹

There is some controversy among epidemiologists about the appropriate role of significance testing.⁸² To the strictest significance testers, any study whose p -

or larger association due solely to random error where, in fact, there is no association.

79. *Allen*, 588 F. Supp at 416–17 (discussing statistical significance and selection of a level of alpha); see also *Sanders*, *supra* note 15, at 343–44 (explaining alpha, beta, and their relationship to sample size).

80. 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 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 Dep't Health & Human Servs.*, No. 92-5089, 1993 U.S. App. LEXIS 29590, at *29, 41 (Fed. Cir. Nov. 15, 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. First, alpha does not address the likelihood that plaintiff's disease was caused by exposure to the agent; the magnitude of the association bears on that question. See *infra* § V. Second, significance testing only bears on whether the magnitude of association found was due to 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 Pharmaceuticals, Inc.*, 911 F.2d 941, 947 (3d Cir. 1990). Fourth, using an alpha of 0.5 would not be equivalent to saying that the probability the association found is real is 50%, and the probability it is a result of random error is 50%. Statistical methodology does not permit assessments of those probabilities. See *Green*, *supra* note 23, at 686; 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; *Allen v. United States*, 588 F. Supp. 247, 417 (D. Utah 1984) ("whether a correlation between a cause and 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 Pharmaceuticals, Inc.*, 959 F.2d 1349, 1357 n.2 (6th Cir.), *cert. denied*, 113 S. Ct. 84 (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.").

81. 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 (alpha frequently set at 95%, though 85% and 90% are also used; selection of the value is "somewhat arbitrary").

82. 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, 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. 884 F.2d at 167.

A number of courts have followed the *Brock* decision or have indicated strong support for significance testing

value does not exceed the level chosen for statistical significance should be rejected as inadequate to disprove the null hypothesis. In the past, authors of a study simply would report whether or not the results were statistically significant, without providing any information about the p -value.⁸³

For others, a statistical device known as a *confidence interval* permits a more refined assessment of appropriate inferences about the association found in an epidemiological study.⁸⁴ The advantage of a confidence interval is that it displays more information than a p -value. What a p -value does not provide is the magnitude of the association found in the study or an indication of how numerically stable that association is. A confidence interval for any study shows the relative risk determined in the study as a point on an axis. It also displays the boundaries of relative risk consistent with the data found in the study based on one or several selected levels of alpha or statistical significance. A sample confidence interval is displayed in Figure 5. The confidence interval represents a study that found a relative risk of 1.5, with boundaries of .8 to 3.4 for alpha equal to .05 and boundaries of 1.1 to 2.2 for alpha equal to .1. Because the boundaries of the confidence limits with alpha set at .05 encompass a relative risk of 1.0, the study

as a screening device. See *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*, 112 S. Ct. 2304 (1992); *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 727 F. Supp. 570, 575 (S.D. Cal. 1989), *aff'd on other grounds*, 951 F.2d 1128 (9th Cir. 1991), *vacated*, 113 S. Ct. 2786 (1993).

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*, 588 F. Supp. at 417, 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) and their use as an alternative to statistical significance in *DeLuca*, 911 F.2d at 948–49. See also *Turpin*, 959 F.2d at 1357 (“The defendant’s claim overstates the persuasive power of these statistical studies. An analysis of this evidence demonstrates that it is possible that Bendectin causes birth defects even though these studies do not detect a significant association.”); *In re 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 epidemiological evidence).

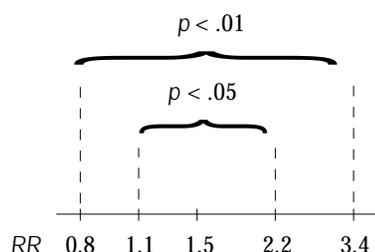
Although the trial court had relied in part on the absence of statistically significant epidemiological studies, the Supreme Court in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 113 S. Ct. 2786 (1993), did not explicitly address the matter. The Court did, however, in identifying factors relevant to the scientific validity of an expert’s methodology, refer to “the known or potential rate of error.” *Id.* at 2797. 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 due to either false matches or false eliminations.

83. Epidemiological studies have become increasingly more statistically sophisticated in their treatment of random error. See *Sanders*, *supra* note 15, at 342 (describing the improved handling and reporting of statistical analysis in studies of Bendectin after 1980).

84. 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 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 epidemiological context is discussed in David H. Kaye & David A. Freedman, Reference Guide on Statistics § IV.A, in this manual.

would not be statistically significant at that level. By contrast, since the confidence boundaries for alpha equal to .1 do not overlap with 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 numerical stability), reflecting the decreased likelihood that the association found in the study would occur if the true association is 1.0.⁸⁵

Figure 5
Confidence Intervals



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 with alpha at .05. The trade-off for reducing false positives is an increase in *false negatives* (also, *beta error* or *type II error*). 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.⁸⁶ The beta for any study can be calculated only based on an alternative hypothesis about a given positive relative risk and

85. Where multiple epidemiological studies are available, a technique known as *meta-analysis* (see *infra* § IV.B.3) can be used to combine the results of the studies to reduce the numerical instability of all. See generally Frederic M. Wolf, *Meta-Analysis: Quantitative Methods for Research Synthesis* (1986). 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 in the epidemiological field. See Zachary B. Gerbarg & Ralph I. Horwitz, *Resolving Conflicting Clinical Trials: Guidelines for Meta-Analysis*, 41 *J. Clin. Epidemiol.* 503 (1988).

In *In re Paoli R.R. Yard PCB Litig.*, 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 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 Pharmaceuticals*, 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 Pharmaceutical 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 control groups in different studies in which some gave the control a placebo and others gave the control an alternative treatment), *cert. denied*, 114 S. Ct. 304 (1993).

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

the level of alpha selected.⁸⁷ 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

The power of a study expresses the likelihood of detecting a postulated level of effect, assuming such an effect exists.⁸⁸ 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 2.0⁸⁹ 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 is low, it has significantly less probative value than a study with similar results but a higher power.⁹⁰

D. What Biases May Have Existed That Would Result in an Erroneous Association?

Systematic error or bias can produce an erroneous association in an epidemiological study.⁹¹ Major sources of bias in the context of planning an epidemiological study were discussed previously in section II. After a study is completed, similar inquiries can be made about whether the study design, data collection, or analysis are flawed and therefore create error. Such an inquiry would be informed by the same concerns described in section II.

Even if one concludes that the findings of a study are statistically stable and that biases have not created significant error, another inquiry remains. An association, as repeatedly noted, does not necessarily mean a causal relationship exists. To make a judgment about causation, a knowledgeable expert must consider the possibility of confounding factors and use several criteria to determine whether an inference of causation is appropriate. These matters are discussed in section IV.

87. See Green, *supra* note 23, at 684–89.

88. For clarification, see *supra* § II.B.2.b and Figure 3.

89. The use of a relative risk of 2.0 for illustrative purposes is because of the legal significance of this magnitude of association. See *infra* § V.

90. See also David H. Kaye & David A. Freedman, Reference Guide on Statistics § IV.B.3.a, in this manual.

91. See *supra* § III.B.

IV. General Causal Association Between Exposure and the Disease

Once an association has been found between exposure to a substance and a disease, researchers consider whether the association reflects a true cause-effect relationship or, alternatively, a spurious finding.⁹² As mentioned in section I, epidemiology cannot prove causation; causation is a judgment issue for epidemiologists and others interpreting the epidemiological data.⁹³

Researchers first look for alternative explanations for the association, such as bias or confounding factors, the latter of which is discussed below. Once this process is completed, researchers consider the guidelines for causation. These guidelines consist of seven inquiries that assist researchers in making a judgment about causation.⁹⁴ As a final step, researchers interpret the data and draw a conclusion about the existence of a cause-effect relationship. 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.⁹⁵

92. 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 familiar “but for” or *sine qua non* test 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.” W. Page Keeton et al., *Prosser & Keeton on Torts* 265 (5th ed. 1984); see also Restatement (Second) of Torts § 432(1) (1965). Epidemiologists use the term to mean that increase in disease among the exposed group would not have occurred in the group 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. See Rothman, *supra* note 32, at 11 (“We can define a cause of a disease as an event, condition or characteristic that plays an essential role in producing an occurrence of the disease.”); *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). Translating the epidemiological concept of cause to the legal question of whether exposure to an agent caused an individual’s disease is addressed *infra* § V.

93. In epidemiology, the practice of drawing inferences about causation is extremely controversial. On one side of this controversy, Professor Kenneth Rothman and his supporters argue that drawing conclusions about causation is not part of science at all, but the domain of public policy. They suggest that scientists should provide policy makers with information but should not advocate a particular interpretation. On the other side of this controversy are more traditional epidemiologists who contend that the researcher is often in the best position to interpret the results, and ought to do so when possible. See Stephan F. Lanes, *Causal Inference Is Not a Matter of Science* (abstract), 122 *Am. J. Epidemiol.* 550 (1985).

94. The guidelines, referred to as “Koch’s postulates” (see *infra* § IV.B), were used first in the field of infectious diseases. See Mervyn Susser, *Causal Thinking in the Health Sciences: Concepts and Strategies in Epidemiology* (1973).

95. In *Cadarian v. Merrell Dow Pharmaceuticals, Inc.*, 745 F. Supp. 409, 412 (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

This section of the reference guide is organized around the following three topics:

1. identification and adjustment for potential confounding factors;
2. application of guidelines for causation; and
3. interpretation of the results.

A. Could a Confounding Factor Be Responsible for the Study Result?⁹⁶

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 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 hair group (those with gray hair tend to be older), old age may be responsible for the association found between hair color and death.⁹⁷ Researchers must separate the relationship between gray hair and risk of death and 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.⁹⁸

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*.⁹⁹ 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? The first 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. It is known that smoking is an important risk factor for cancer of the pancreas. It also is known 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

support an inference of causation in the absence of independent confirmatory studies. The court did not address the question of whether the degree of certainty employed by epidemiologists before making a conclusion of cause was consistent with the legal standard. See *DeLuca v. Merrell Dow Pharmaceuticals, 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 Pharmaceutical Corp.*, 788 F.2d 741, 745 (11th Cir.), *cert. denied*, 479 U.S. 950 (1986).

96. 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).

97. This example is drawn from Kahn, *supra* note 36, at 63.

98. Similarly, a finding of no association may be erroneous because a confounding factor with a protective effect is differentially associated with those exposed to the agent. One example of a confounding factor with a protective effect is vaccination.

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

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 consider the possible confounding role of cigarettes, MacMahon examined smokers and nonsmokers separately to determine whether the relationship between coffee and cancer of the pancreas held in both groups. When smoking was held constant, he still found an increasing risk of pancreatic cancer with increasing consumption of coffee, particularly in women.

The main problem in many observational studies such as MacMahon's is that the individuals are not assigned randomly to the exposed cohort and the control group.¹⁰⁰ Instead, individuals self-select themselves for that exposure (or in many studies someone else selects them), a feature of virtually all observational human population studies without *randomization*. 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 have been "selected"—by themselves or by others—based on residence, socioeconomic status, age, or other factors.¹⁰¹ These other selection factors may be causing the disease, but because of the selection an apparent (yet false) association of the disease with exposure to the agent may appear.

Confounding factors that are known in advance can be controlled during the study design and through study group selection. Unanticipated confounding factors that can be identified can sometimes be controlled during data analysis if data are gathered about them. There is always a risk, however, that an undiscovered confounding factor is responsible for a study's findings.¹⁰²

1. What techniques, if any, were used to identify confounding factors?

Care in the design of a research project (e.g., methods to select the subjects, diagnose disease, and assess exposure) can prevent confounding. To identify potential confounding factors, the researcher must assess a range of factors that could influence risk. This procedure often involves complex statistical manipulation to compare the overall risk of exposure with the risk when identified potential confounding factors have been removed from the calculation.

Using MacMahon's study as an example, the researcher would test 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

100. Randomization assumes 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. See the Glossary of Terms for additional comments on randomization and confounding.

101. 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).

102. Rothman, *supra* note 32, at 125.

drinkers. If the risk is 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).

2. What techniques, if any, were used to control 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 of the confounding variable. Statistical methods then can be applied to combine the different results of 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). By comparing the different rates of pancreatic cancer for people in each group who drink the same amount of coffee, 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 the confounding factor through mathematical modeling. *Models* are developed to describe the simultaneous effect of exposure and confounding factors on the increase in risk. This technique relies on building a series of mathematical models to predict who will get the disease.¹⁰³ For instance, MacMahon might have begun a multivariate analysis with a simple model to determine how well the individual's daily intake of coffee predicts whether he or she will contract pancreatic cancer. In the next model, he could add the number of years the person had been a coffee drinker. If the second model better predicts who would contract cancer, MacMahon would continue to create more complex models (including variables such as age, gender, and ethnic group) until he found a model that best predicts who will contract cancer.

If the association between exposure and disease remains after completing the assessment and adjustment for confounding factors, the researcher applies the guidelines described in section IV.B to determine whether an inference of causation is warranted.

B. Overall, Does Application of the Guidelines for Causation Support a Finding of Causation?

Seven factors should be considered when an epidemiologist determines whether

¹⁰³ For a more complete discussion, see Daniel L. Rubinfield, Reference Guide on Multiple Regression, in this manual.

the association between an agent and a disease is causal.¹⁰⁴ These factors guide the epidemiologist in making a judgment about causation. They are

1. strength of the association;
2. temporal relationship;
3. consistency of the association;
4. biologic plausibility (coherence with existing knowledge);
5. consideration of alternative explanations;¹⁰⁵
6. specificity of the association; and
7. dose-response relationship.

These guidelines, known as Koch's postulates, were proposed first about 100 years ago by two infectious disease researchers, Koch and Henle.¹⁰⁶ Each factor is considered in the following subsections.

1. How strong is the association between the exposure and disease?¹⁰⁷

The relative risk is one of the cornerstones for causal inferences.¹⁰⁸ Relative risk measures the strength of the association. The higher the relative risk, the greater the likelihood that the relationship is causal.¹⁰⁹ For cigarette smoking, for example, the estimated relative risk for lung cancer is very high, about 10.¹¹⁰ That is, the risk of lung cancer in smokers is approximately nine to ten times the risk in nonsmokers.

A relative risk of 9 to 10, as seen with smoking and lung cancer, is so high that it is extremely difficult to imagine any kind of error in the study that would have produced it. The higher the relative risk, the stronger the association, and the more likely an epidemiologist will consider it causal. Although lower relative risks can reflect causality, the epidemiologist will scrutinize the association more closely.

Attributable risk, another measure of excess risk, is particularly important in the legal arena, because it measures the excess risk caused by exposure to the

104. See *Smith v. Ortho Pharmaceutical Corp.*, 770 F. Supp. 1561, 1575–76 (N.D. Ga. 1991).

105. See *supra* note 94 and accompanying text.

106. The two factors, dose-response relationship and specificity of the association, are not always used. See *infra* note 119 and accompanying text.

107. 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 the individual plaintiff's injury. See *infra* § V.

108. See *supra* § III.A.1.

109. The reason that a higher relative risk is more likely to indicate a true causal relationship is because such a strong effect is unlikely to be the result of bias or random sampling error. Findings of small relative risks are much more susceptible to these errors. 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 are rarer phenomena than stronger effects. See *Green*, *supra* note 23, 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*, *supra* note 32, at 12–13.

110. See *Doll & Hill*, *supra* note 6.

agent.¹¹¹ For example, if a group of individuals is exposed to PCBs and has a high risk of cancer, attributable risk permits the epidemiologist to subtract the background risk of disease from the exposed group's total risk of disease. In doing so, the epidemiologist measures the increased risk of disease that can be attributed to a specific exposure, which can then be used to determine the benefit that would be gained by eliminating a particular exposure.¹¹²

2. 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.¹¹³ If the exposure occurs after the disease develops, it cannot cause the disease.

3. Is the association consistent with other research?¹¹⁴

The need to replicate research findings permeates most fields of science. In epidemiology, research findings often are replicated in different populations.¹¹⁵ Consistency in these findings is an extremely important factor in making a judgment about causation. Different studies that examine the same exposure-disease relationship should yield similar results. Any inconsistencies signal a need to question whether the relationship is causal.

Meta-analysis is an analytic technique that allows epidemiologists to combine the results of several research studies to better understand the relationship between exposure to an agent and a disease.¹¹⁶ The combined data are analyzed to determine if they render different results from those in the individual studies performed with smaller sample sizes.¹¹⁷ Particular concern must be paid to the

111. Risk is not zero among the control group (those not exposed) when there are other causal chains that cause the disease that do not require exposure to the agent. For example, a proportion of 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 Env'tl. 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 122.

112. The benefit gained by eliminating a particular exposure would be equivalent to the amount of disease that could be prevented by eliminating that exposure. See *supra* § III.A for an example of how to calculate this amount.

113. 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. . . .").

114. 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.

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

116. See *In re Paoli R.R. Yard PCB Litig.*, 916 F.2d 829, 856-57 (3d Cir. 1990), cert. denied, 499 U.S. 961 (1991); *Hines v. Consolidated Rail Corp.*, 926 F.2d 262, 273 (3d Cir. 1991).

117. See *supra* note 85.

propriety of combining different study populations and to the appropriate inferences to be drawn from the meta-analysis.

4. Is the association biologically plausible (consistent with existing knowledge)?¹¹⁸

Biological plausibility is not a simple criterion to use. When an association is biologically plausible, the plausibility is appealing and provides supporting evidence. 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 epidemiological 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 others.

5. Have alternative explanations been ruled out?

Alternative explanations and confounding factors should be examined and ruled out to avoid reaching an erroneous conclusion. However, it is never possible to rule out every alternative explanation. Epidemiology cannot prove causation. It is an inference for the scientist to make and usually is not made lightly.

The last two factors, specificity of the association and dose-response relationship, differ in significant ways from the five factors mentioned above. Although the presence of specificity and dose-response strengthens the inference of causation, the absence of either does not weaken the inference. Epidemiologists have begun to question the use of these two factors as guidelines for causation in non-infectious diseases.¹¹⁹

6. Does the association exhibit specificity?

An association exhibits specificity if the exposure is associated only with a single disease or type of disease. As mentioned above, epidemiologists no longer require that the effect of exposure to an agent be specific for a single disease. For example, cigarette manufacturers have long claimed that since cigarettes have

118. 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*, in this manual.

119. Koch's postulates were originally formulated for determining causation of infectious diseases. Specificity and dose-response remain important factors in infectious disease epidemiology. *See supra* § IV.B and note 106.

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. The scientific bases that have undermined the guideline include the following: (1) Human cells and tissues share many common features. They all have a basic structure, including nuclei, DNA, and other characteristics. There is every reason to expect that a certain agent will act on certain cellular components and structures even if they are in different tissues and different organs; and (2) Tobacco and cigarette smoke are not single agents but mixtures of harmful agents. Smoking represents exposure to multiple agents and specificity would not be expected. However, most known teratogens cause a specific birth defect or a related pattern of birth defects.

7. Is there a dose-response relationship?

A dose-response relationship assumes that the more intense the exposure, the greater the risk of disease. However, the researcher may not observe a dose-response relationship when there is a *threshold phenomenon* (i.e., a low dose exposure may not cause disease until the exposure exceeds a certain dose).¹²⁰ Evidence of a dose-response relationship strengthens the conclusion that the relationship between an agent and disease is causal; however, a dose-response relationship is not necessary to infer causation.

C. What Type of Causal Association Has Been Demonstrated Between Exposure and Disease?

Assuming an association is not due to confounding factors and that the epidemiologist has decided that the scientific findings overwhelmingly support an inference of causation, the epidemiologist next determines which type of causal relationship exists between the agent and the disease in the exposed population. Epidemiologists divide causes into four categories (see Figure 6). It should be noted that the terms applied to the four categories of causation are not consistent

120. The question of 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. Dep't 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).

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 question of 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).

with legal terminology. Nevertheless, these terms may be useful in understanding them when they appear in a published study or when used by an epidemiologist.¹²¹

Figure 6
Four Categories of Causation

1. Necessary and sufficient (occurs rarely)

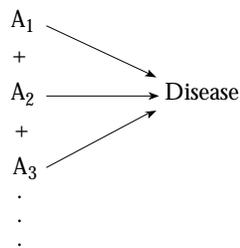
Factor A \longrightarrow Disease

2. Necessary but not sufficient

$A_1 + A_2 + A_3 \dots \longrightarrow$ Disease

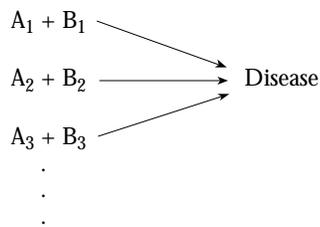
(Causal chain may also involve a specific temporal sequence)

3. Sufficient but not necessary



4. Neither necessary nor sufficient

(probably true for most of the diseases we study)



1. Exposure to an agent may be a *necessary and sufficient cause* of the disease. This type of causal relationship assumes that the disease will not result unless an individual is exposed. Nothing but the agent is needed to cause the disease.
2. Exposure can be *necessary but not a sufficient cause* of the disease. In

121. See Mervyn Susser, *What Is a Cause and How Do We Know One? A Grammar for Pragmatic Epidemiology*, 133 *Am. J. Epidemiol.* 635, 637 (1991).

this causal relationship the disease will not result unless an individual is exposed, but exposure in and of itself is not enough to cause the disease.

3. An exposure may be a *sufficient but not necessary cause* of the disease when the disease occurs not only in the presence of exposure but also in the presence of exposures to other agents. Leukemia is an example of this relationship; exposure to radiation or benzene can result in the occurrence of disease.
4. The last possibility is that exposure is *neither a necessary nor sufficient cause* of the disease. This takes place when the disease occurs in the absence of exposure and does not always occur in its presence. This complicated relationship is probably the one that most faithfully represents the causal relationships in the majority of diseases encountered. The disease can occur through a variety of combinations of different exposures.

V. The Role of Epidemiology in Proving Individual 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.¹²² 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 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 plaintiff's disease.¹²³

Nevertheless, the specific causation issue is a necessary 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 epidemiological evidence plays in answering that question. This question is not a question about which an epidemiologist would have any expertise to contribute. Rather it is a legal question with which a number of courts have grappled. An explanation of how these courts have resolved this question follows.

There are two legal issues that 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 epidemiological study that is suffi-

122. See *DeLuca v. Merrell Dow Pharmaceuticals, 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 Pharmaceutical 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 sometimes referred to as a "signature disease," 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).

123. *Cf. In re "Agent Orange" Prod. Liab. Litig.*, 597 F. Supp. at 780.

ciently rigorous to justify a conclusion that it is scientifically valid should be admissible,¹²⁴ as it tends to make an issue in dispute more or less likely.¹²⁵

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.”¹²⁶ The relative risk from an epidemiological study can be adapted to this 50% plus standard to yield a probability or likelihood that an agent caused an individual’s disease.¹²⁷ The threshold for concluding that an agent was more likely the cause of a disease than not 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 implies a 50% likelihood that an exposed individual’s disease was caused by the agent. A rela-

124. See *DeLuca*, 911 F.2d at 958; cf. *Kehm v. Procter & Gamble Co.*, 580 F. Supp. 890, 902 (N.D. Iowa 1982) (“These [epidemiological] studies were highly probative on the issue of causation—they all concluded that an association between tampon use and menstrually related TSS [toxic shock syndrome] cases exists.”), *aff’d 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 1); 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 epidemiological studies were admissible despite criticism of the methodology used in the studies. The court held that the claims of bias went to the weight rather than the admissibility of the studies. 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*, 112 S. Ct. 1280 (1992).

125. 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. Exclusion of an otherwise relevant epidemiological study on Rule 403 grounds is unlikely.

In *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 113 S. Ct. 2786, 2796 (1993), the Court invoked the concept of “fit,” which addresses the relationship of an expert’s scientific opinion with the facts of the case and the issues in dispute. In a toxic substance case in which cause in fact is disputed, an epidemiological 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 almost surely 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.

126. 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).

127. 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 requires a shift from a frequentist approach, which involves sampling or frequency data from an empirical test, to a subjective probability about a discrete event. Thus, a frequentist might assert, after conducting a sampling test, that 60% of the balls in an opaque container are blue. The same frequentist would resist the statement: “The probability that a single ball removed from the box and hidden behind a screen is blue is 60%.” The ball is either blue or not, and no frequentist data would permit the latter statement. “[T]here is no logically rigorous definition of what a statement of probability means with reference to an individual instance. . . .” Lee Loevinger, *On Logic and Sociology*, 32 *Jurimetrics J.* 527, 530 (1992); see also Steve Gold, 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 80, at 54–62; David H. Kaye & David A. Freedman, *Reference Guide on Statistics* § IV.B, in this manual.

tive 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. A substantial number of courts in a variety of toxic substances cases have accepted this reasoning.¹²⁸

An alternative, yet similar means to address probabilities in individual cases is by use of the attributable proportion of risk parameter.¹²⁹ The attributable risk is a measurement of the excess risk that can be attributed to an agent, above and beyond the background risk due to other causes. When the attributable risk exceeds 50% (equivalent to a relative risk greater than 2.0), this logically might be converted to a belief that the agent was more likely than not the cause of the plaintiff's disease.

The discussion above assumes that the only evidence bearing on cause in fact is epidemiological. Such an assumption is unlikely, and a variety of additional pieces of evidence, although less quantifiable, affect a fact finder's assessment. Biases in the epidemiological studies might justify a conclusion that the real magnitude of increased risk is greater or lower than that revealed in the studies. The dose to which the plaintiff was exposed may be greater or lesser than those in the epidemiological study, thereby requiring some extrapolation.¹³⁰ In addition, there may be factors peculiar to the plaintiff—excess exposure to another known cause, pathological mechanism,¹³¹ family history of disease, or conflicting diagnoses—that modify any probability based solely on the available epidemiological evidence.¹³²

128. See *DeLuca v. Merrell Dow Pharmaceuticals, Inc.*, 911 F.2d 941, 958–59 (3d Cir. 1990) (Bendectin allegedly caused limb reduction birth defects); *In re Joint E. & S. Dists. Asbestos Litig.*, 758 F. Supp. 199, 203 (S.D.N.Y. 1991) (asbestos allegedly caused colon cancer), *rev'd*, 964 F.2d 92 (2d Cir. 1992) (relative risk less than 2.0 may still be sufficient to prove 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); *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"). *But cf. In re Fibreboard Corp.*, 893 F.2d 706, 711–12 (5th Cir. 1990) (The court disapproved a trial in which several representative 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 general causation, which ignores any proof particularistic to the individual plaintiff, could not substitute under Texas law for cause in fact.).

129. See *supra* § III.A.3.

130. See *supra* § IV.B.5; 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. Dists. Asbestos Litig.*, 774 F. Supp. 113, 115 (S.D.N.Y. 1991) (discussing different relative risks associated with different dose levels), *rev'd on other grounds*, 964 F.2d 92 (2d Cir. 1992).

131. See *Tobin v. Astra Pharmaceutical Prods., Inc.*, 993 F.2d 528 (6th Cir.) (plaintiff's expert relied predominantly on pathogenic evidence), *cert. denied*, 114 S. Ct. 304 (1993).

132. An example of a judge sitting as fact finder and considering individualistic 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

This additional evidence bearing on 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 epidemiological evidence.¹³³ For example, genetics might be known to be responsible for 50% of the incidence of a disease. 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.¹³⁴

Manko v. United States, 636 F. Supp. 1419, 1437 (W.D. Mo. 1986), *aff'd*, 830 F.2d 831 (8th Cir. 1987).

133. 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. Dists. 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).

134. 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 double may not be required to pay 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 ed., 1988); 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).

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 have a single agent, a number of independent alternative agents, or a complex of two or more factors whose combined presence is necessary for the development of the disease (e.g., a virus is the agent of measles).

Alpha. The level of statistical significance chosen by the researcher to determine if any association found in the study is sufficiently unlikely to have occurred by chance (due to 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. Alpha error, also called type I error, occurs when the 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 and experimental groups, 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 dependence 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 (APR) . 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).

Background Risk of Disease . Background risk of disease (or background rate of disease) is the amount of disease in a population that occurs in individuals who have 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 . Beta error, also called type II error or false negative, occurs when the 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 and experimental groups, 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. 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 biological marker is 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.

Biological Plausibility. This factor considers 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 the disease (or other outcome variable) and a suitable control (comparison, reference) group of persons without the disease. Such a study is called 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 that is a necessary element of a set of other events that produce an outcome, such as a disease. Thus, a cause may be thought of as a necessary link in some causal chain that results in an outcome of interest.

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 a factor or factors 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 within which the results of a study sample would be likely to fall if the study were repeated numerous times. Thus, if a p -value of .05 is selected, a confidence interval would indicate the range of relative risk values that would result 95% of the time if the study were repeated. 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. A confounding factor is both a risk factor for the disease and 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 (identified as a rule before a study is begun) comprising individuals who have not been exposed to the disease, intervention, procedure, or other variable whose influence is being studied. In statistics, control procedures try to filter out the effects of confounding variables on nonexperimental data, typically by “adjusting” through statistical procedures (like multiple regression).

Dose. Dose generally refers to the intensity or magnitude of exposure multiplied by the duration of exposure.

Dose-Response Relationship. A relationship in which a change in amount, intensity, or duration of exposure is associated with a change—either an increase or a decrease—in risk of disease.

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 in-

fluences the individual's chances of contracting or dying of heart disease.

Effect Size. The effect size, or magnitude of the increased risk in disease, is best thought of as the amount of disease that is caused by exposure to a toxic substance.

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

Error. Random error (sampling error) is that due to chance when the result obtained in the sample differs from the result that would be obtained if the entire population (universe) were studied. Two varieties of sampling error are type I error, or alpha error, and type II error, or beta error.

When hypotheses testing is used, rejecting a null hypothesis when it is actually true is called type I error. Failing to reject a null hypothesis when it is incorrect is called type II error.

Etiologic Factor. An agent that plays a role in causing a disease.

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.

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. 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.

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

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 better suited to pooling results from randomly controlled experimental studies, but if carefully performed, it also may be used for

observational studies.

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 used to examine consequences of various actions, or (2) a representation of a country's situation through an "average region" with characteristics resembling those of the whole country.

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 factors or variables.

Null Hypothesis. At the outset of any observational or experimental study, the researcher must state a principle or proposition that will be tested in the study. In epidemiology, this principle typically addresses the existence of a causal relation between an agent and a disease. Most often, the null hypothesis is a statement that Agent A does not cause Disease D. The results of the study may justify a conclusion that the null hypothesis 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 does not justify a conclusion that the null hypothesis has been proved.

Observational Study. An observational study is an epidemiological study in situations where 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 two odds. For most purposes the odds ratio from a case-control study is quite similar to a risk ratio from a cohort study.

P (Probability), p-Value. The *p*-value is the probability of getting a value of the test statistic 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) or by the symbol for less than (<) and a decimal notation such as .01 or .05, is a statement of the probability that the difference observed could have occurred by chance.

Investigators may arbitrarily set their significance levels, but in most biomedical and epidemiological work, a study result whose probability value

is less than 5% ($p < .05$) or less than 1% ($p < .01$) is considered sufficiently unlikely to have occurred by chance to justify the designation statistically significant.

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.

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 each group is compared. See Cohort Study.

Random. The term implies that an event is governed by chance. See Randomization.

Randomization. Allocation of individuals to groups (e.g., for experimental and control regimens) by chance. Within the limits of chance variation, randomization should make the control and experimental groups similar at the start of an investigation and ensure that personal judgment and prejudices of the investigator do not influence allocation.

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 be used where the exposure is known to cause harm (e.g., cigarette smoking).

Relative Risk (RR). The ratio of the risk of disease or death among the exposed 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.

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 and may be representative or nonrepresentative.

Sample Size. The number of subjects who participate in a study.

Secular Trend Study. Also, time-line study. This type of study 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 coro-

nary heart disease mortality in the United States in the past fifty years.

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, then 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).

Statistical Significance. This term is 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, socioeconomic status, and so forth. The effect of confounding variables may be controlled by stratifying 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 population may be divided 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.

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.

Toxicology. The science of the nature and effects of poisons, their detection, and the treatment of their effects.

Toxic Substance. A substance that is poisonous.

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

Type I Error. See Alpha Error and Error.

Type II Error. See Beta Error and Error.

Validity. The degree to which a measurement measures what it purports to measure.

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 ed., 1988).
- Joseph L. Fleiss, *Statistical Methods for Rates and Proportions* (1981).
- Harold A. Kahn, *An Introduction to Epidemiologic Methods* (1983).
- Harold A. Kahn & Christopher T. Sempos, *Statistical Methods in Epidemiology* (1989).
- Abraham M. Lilienfeld & David E. Lilienfeld, *Foundations of Epidemiology* (1980).
- Judith S. Mausner & Anita K. Bahn, *Epidemiology: An Introductory Text* (1974).
- Kenneth J. Rothman, *Modern Epidemiology* (1986).
- 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 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).

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

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).

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

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).

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).

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).

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

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).