Understanding Epidemiology and Its Use in Drug and Medical Device Litigation

Defense counsel must be prepared to combat expert testimony that is inconsistent with the sound use of epidemiological data

By Bruce R. Parker

IADC member Bruce R. Parker is a partner in the Baltimore firm of Goodell, DeVries, Leech & Gray, where his practice is concentrated in the areas of product liability. He is a graduate of Johns Hopkins University (1975) and the Columbus School of Law of the Catholic University of America (1978).

OVER the last 20 years, there has been an explosion in toxic tort industry-wide litigation. Drugs and medical devices have not been exempt from the epidemic of lawsuits. Notable examples include Bendectin, DES, L-Tryptophan, the Dalkon shield, tampons and silicone breast implants. Among the explanations for the explosion is the deference courts have given to plaintiffs' experts who have been permitted to give causation opinions based on nothing more than personal speculation dressed up to sound like scientific evidence. Few decisions before *Daubert v. Merrell Dow Pharmaceuticals Inc.*[1] recognized the critical role that epidemiology plays in the scientific method for ascertaining causation.

Daubert has had many effects, not the least of which is to compel trial courts to examine critically the validity of the scientific methodology that purportedly supports an expert's testimony. This means that defense counsel need to understand the role epidemiology plays in the scientific search for the cause of disease and how to differentiate good from weak epidemiological evidence.

The goal of this article is to provide an overview of basic epidemiological principles and to discuss how such evidence is used and misused in trials. It is not an exhaustive treatise on epidemiological study designs and statistical analysis. However, it addresses the major issues raised in defending a drug or medical device manufacturer against claims that the product caused systemic disease. The article contains a general discussion on what epidemiology is and the role it plays in the scientific investigation of the cause of disease, discusses how an epidemiological study is performed and the differences among epidemiological studies, addresses statistical issues raised by epidemiological studies and questions to consider when interpreting the reliability and validity of epidemiological data, discusses how and under what circumstances epidemiological data can be used to reach conclusions regarding causation, considers how courts have used epidemiology, and offers practice pointers on cross examining plaintiffs' epidemiologist and presenting epidemiological testimony in the defense case.

WHAT IS EPIDEMIOLOGY?

Epidemiology can be broadly defined as "the study of the distribution and determinants of health-related states or events in specified populations and the application of this study to control of health problems."[2] The objectives of epidemiology are to: (1) identify what factors increase the risk of disease; (2) determine how many

1. 509 U.S. 579 (1993).
2. J. LAST, A DICTIONARY OF EPIDEMIOLOGY (2d ed. 1988).

new cases of a disease develop in a defined period (incidence) and how many total cases of the disease exist at specified points in time (prevalence); (3) obtain an understanding of how a disease develops in terms of time and prognosis; (4) evaluate the efficacy of new therapies, such as surgery and drugs;[3] and (5) assist in formulating regulatory decisions regarding environmental issues that affect the rate at which diseases develop and persist.[4]

All epidemiological studies, regardless of whether their goal is to study risk factors of disease or efficacy of treatment modalities, share a common feature. What distinguishes an epidemiological study from a case report or case series is the inclusion of a control group that permits the investigator (assuming the study is done correctly) to draw reasonable conclusions with respect to the effect, if any, of the exposure or treatment being studied. Proper scientific methodology, as required by *Daubert*, does not permit conclusions to be drawn regarding a causal relationship between exposure and disease or the efficacy of treatment based on case reports or case series.[5]

A number of examples can be used to demonstrate this point. For example, many case series of patients with low back pain who had undergone MRI concluded that the majority of the patients with back pain had disc abnormalities, leading some practitioners to conclude that disc abnormalities were a cause of low back pain. However, later studies using a controlled group of patients who did not have back pain revealed similar rates of disc abnormalities. Such evidence does not support a conclusion that disc abnormalities as revealed by MRI are the cause of low back pain.[6]

Unlike animal studies and in vitro cell testing, an epidemiological study is conducted on human populations and is therefore uniquely capable of quantifying the exposure-disease relationship.[7] Epidemiology is not saddled with the difficulties of extrapolating the data to human beings, as are the other forms of scientific evidence. For example, the best animal experiment is always subject to the question of its applicability to human beings. Similarly, the results of laboratory (in vitro) studies on human cells are not necessarily good predictors of what happens in the body since the laboratory environment is often dissimilar to the in vivo condition.

For these reasons, as discussed later, courts are increasingly viewing epidemiological evidence as the most probative scientific evidence of causation.[8] Although a well-designed, properly performed epidemiological study provides the most probative scientific evidence on causation, there are limitations to the results. As explained

3. Using epidemiological tools to study the efficacy of treatment and preventative measures is called clinical epidemiology, which is defined as the "science of making predictions about individual patients by counting clinical events in similar patients, using strong scientific methods for studies of groups of patients to ensure that the predictions are accurate. The purpose of clinical epidemiology is to develop and apply methods of clinical observation that will lead to valid conclusions by avoiding being misled by systematic error and chance." ROBERT H. FLETCHER et al., CLINICAL EEPIDEMIOLOGY 2 (1996).

4. *See* LEON GORDIS, EPIDEMIOLOGY 3 (1996).

5. *See* Fletcher, *supra* note 3, at 208-12, which provides a comprehensive discussion of the role of case reports and case series and the inappropriateness of relying on such data for reading conclusions on causation or altering clinical practice.

6. *Id.* at 212.

7. CHARLES H. HENNEKENS & JULIE E. BURING, EPIDEMIOLOGY IN MEDICINE 12-13 (1987).

8. *See, e.g.*, Richardson v. Richardson-Merrell Inc., 857 F.2d 823, 830 (D.C. Cir. 1988), *cert. denied*, 493 U.S. 882 (1989); REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 126 (Federal Judicial Center 1994). Although courts have been arguably slow to appreciate the unique probative nature of epidemiological data, its critical role in scientific investigations can be traced back several hundred years. However, the first modern application of epidemiology began in the 1850s with the discovery by John Snow, after he had undertaken an epidemiological investigation, that outbreaks of cholera in London were caused by contaminated water supplies in various districts of London. Fletcher, *supra* note 3, at 6.

later, a study may reveal a statistically significant association between exposure and disease. Even a strong statistical association between exposure and disease, however, does not in and of itself prove that the exposure is capable of causing disease generally. It follows that epidemiological data of a strong association between exposure and disease provides no evidence that in a particular individual the exposure has caused the plaintiff's injury. Finally, although most epidemiological studies generate data that reflect the degree, if any, of the statistical association between a suspected risk factor and disease, epidemiological studies are equally capable of revealing whether a risk factor has the opposite effect of providing a protective mechanism against the occurrence of disease.

HOW ARE EPIDEMIOLOGICAL STUDIES DONE?

A. Observation Versus Experimental Studies

A controlled epidemiological study can be either experimental or observational. The primary difference between the two is the role of the investigator. In an experimental study, which is also called a clinical study, the investigator randomly and blindly selects which of the subjects will be exposed to the agent being studied. In well-conducted experimental studies, the subjects are not told ("blind") whether they have been exposed (*e.g.*, been given the new drug or placebo). Because it would be unethical for investigators to expose people to substances known to be harmful, experimental epidemiology is limited to evaluating new medical treatments that are thought to be beneficial. Because people are randomly exposed in a clinical study, such a study provides the best scientific data on the association, if any, between exposure and the outcome being studied ("endpoint").

As explained later, an association revealed in an epidemiological study may be the result of random chance. If study subjects are exposed by random selection, then it is less likely that chance has produced the results.[9]

In contrast, in an observational study the investigator does not select which subjects are to be exposed. Rather, the investigator selects subjects who have already been exposed and observes which of the exposed and controls develops an illness.

B. Types of Observational Studies

Observational studies are either cohort, case-controlled or nested case-controlled in their study design. The variables that make one person different from another, such as age, genetics, lifestyle, race, sex and exposure duration cannot be controlled directly by investigators in an observational study. Consequently, the effect, if any, of such variables on the rate of disease are controlled in designing a study and by using appropriate statistical tests to analyze the data.[10] The quality of epidemiological data is directly correlated to the design of the study. Studies that are better designed to control for variables that can affect the outcome of the study produce more reliable data on whether a true association exist between exposure and disease.

Whether a cohort, case-controlled or nested case-controlled type of study is performed depends on a variety of factors, including cost, availability of exposure data and whether the disease is rare or common. In general, however, all three studies are capable of generating useful epidemiological data on the relationship between exposure and disease.

1. Cohort Studies

A cohort is a group that is traced over time. A cohort study is one in which groups of exposed and non-exposed subjects are followed over time. Observations

9. *See generally* Hennekens & Buring, *supra* note 7, at 26.

10. Reference Manual, *supra* note 8, at 129.

are made as to how many subjects in both groups develop the disease of interest.

a. Prospective/Concurrent Cohort Study

In a prospective cohort study, investigators begin by identifying a population of people who have been exposed to the agent (such as a chemical) being studied. The control population is comprised of people who, in all important respects, are similar to the exposed population except for the fact that they have not been exposed. The investigators, either by periodic physical examination, questionnaires or other means, will follow both populations over time until the pre-determined period of study has elapsed. At the conclusion of the study, an analysis is conducted of the proportion of exposed subjects who develop one or more diseases of interest. The rate of disease in the exposed population is then compared to the rate of disease in the non-exposed population to determine if the exposure is associated with an increased rate of disease.

The term "bias," discussed more fully later, for present purposes can be defined as any systematic error in performing a study that leads to erroneous data. For example, a study whose design includes people involved in litigation and aware that the study results may have a direct impact on the litigation may be subject to litigation bias. If that bias exists, it raises questions regarding the reliability of the epidemiological data. Although prospective/concurrent cohort studies are free of some of the biases that can affect a retrospective or historical cohort study, most cohort studies, particularly those studying chronic diseases, are retrospective.

Many chronic diseases, unlike infectious diseases, tend to be associated with reasonably long latency periods following the onset of exposure. Consequently, a prospective cohort study of a latent disease would have to be conducted for the length of the expected latency period, which in some cases could be decades. Such studies could "outlive" the investigators. As a result, the vast majority of cohort studies of chronic diseases are retrospective.

b. Retrospective or Historical Cohort Studies

Similar to a prospective cohort study, in a retrospective cohort study, investigators identify an exposed and nonexposed population. Also, the design of the study must be free of the biases and the confounding factors described later.

After investigators have determined the latency period to be studied, medical records generated at the beginning of the latency period are reviewed. All pertinent records generated throughout the retrospective or historical latency period for both population groups are examined to determine the proportion in each group who developed the diseases of interest.[11]

The benefit of a retrospective cohort study is that it can be done reasonably quickly at considerably less cost than a prospective study. The primary disadvantage, however, that counsel should carefully remember when reviewing such data is that it is susceptible to more forms of bias than a prospective cohort study.

c. Benefits and Disadvantages of Cohort Studies

A principal feature of a cohort study, whether prospective or retrospective, is that it is capable of determining the rate or incidence of a disease. Incidence is defined as "the number of new cases of a disease that occur during a specified period of time in a population at risk for developing the disease." Determining the number of new cases of disease over a specified period of time indicates, among other things, the risk

11. Historical cohort studies are often called both retrospective and prospective, because the process involves beginning the study in the past (retrospective) and reviewing the records to the current date (prospective).

of developing the disease.[12]

Assuming that a causal link exists between exposure and disease, a cohort study can provide data on the latency period of the disease. The incidence of disease is calculated as follows:

$$\frac{\text{Number of new cases of a disease which develop in a defined time period}}{\text{Number of persons who are studied who are at risk of developing the disease during the defined period of time}} \times 100 = \text{Incidence per 100}$$

As another benefit, a cohort design allows investigators to observe a number of different diseases which may develop in the exposed and controls. Statistical analysis can be done on whatever diseases develop (or in epidemiological jargon, the "endpoints") to determine if any are a risk factor.[13]

A major disadvantage of the cohort study is that scientists must study large numbers of people in order to detect an association with a rare disease (*e.g.*, has a low incidence rate). For example, assume that the disease studied has an annual incident rate of 1 per 10,000 people. A cohort study containing only 1,000 exposed people would not be expected to generate any cases of the disease, even if no association existed, unless the study were conducted for 10 years.

Scientists are better able to study rare exposures in a cohort study than in a case-control study. Assume, for example, that a suspected association exists between uranium and cancer. Since uranium exposure is relatively rare, it would be difficult to find an association even if one existed between uranium and cancer by using a case-control study, because that study would require identifying patients with cancer and then determining if any had been exposed to the uranium. A cohort study would allow an investigator to select a group of uranium industrial workers and to follow the group over a period of time to determine the proportion, if any, that developed cancer.[14]

Cohort studies also are subject to the same design problems found in case-control studies regarding bias and confounding variables.

2. Case-Control Studies

A case-control study, often described as a retrospective or prevalence study, occurs when investigators select people ("cases") who have already developed the disease endpoint. A control group is assembled whose constituents are matched (except for disease) to the test groups with respect to variables (such as age and sex) that could independently affect either the exposure or the onset of disease. These variable are called confounding factors. The investigators determine through patient interviews, questionnaires or medical records the number of cases and controls who have a history of exposure.

Although case-control studies are generally retrospective, their design permits them to be retrospective or prospective. Therefore, the primary feature that differentiates a cohort from a case-control study is whether the study begins with a group that already has the disease (case-control) or a group that has been exposed but has not yet developed the disease of interest (cohort).

Obviously, those in the exposed group who ultimately are determined to have developed the disease had the disease at the initiation of the study. In a sense, the situation is no different from a case-control

---

12. Gordis, *supra* note 4, at 31.

13. As more comparisons are performed, the probability increases of finding an association that in fact does not exist. Such a finding is referred to as a false positive result. The problems of multiple comparisons are discussed later.

14. Hennekens & Buring, *supra* note 7, at 156-57.

study. The fundamental difference is that the test population of a cohort study is determined exclusively on the basis of its exposure status. If subjects are included within the exposed cohort with knowledge of their disease status, the study is subject to considerable selection bias and the quality of the data is open to question.

a. Prevalence

Case-control studies provide data on the prevalence of diseases. Prevalence is that portion of a population that has the disease at a defined point in time. Prevalence is computed by this formula:

$$\frac{\text{Number of disease cases in population at specified time}}{\text{Number of persons in population}} \times 100 = \text{prevalence per 100}$$

Prevalence data provides information only as to the number of cases existing at a specific point in time but no information on the risk of developing the disease, because it does not measure the rate at which the disease develops.[15] The distinction between incidence and prevalence and their significance is frequently misstated by so-called experts. The relationship between incidence and prevalence is reflected in this equation:

$$\text{Prevalence} = \text{incidence} \times \text{duration of disease}$$

For example, assume 200 people are screened for hepatitis, with 100 located in town A and the other 100 in town B. Blood tests are positive in 50 of the people from town A and 20 from town B. Many self-declared experts will tell jurors that the risk of hepatitis is greater in town A because a greater percentage of people in town A had hepatitis.

Such testimony, however, is erroneous because the information does not shed light on the duration of the disease, which could be affected by independent variables. For example, if there was better health care in A, the people with hepatitis could be living longer than those in B, and therefore the prevalence of the disease could be greater without there being any difference in the incidence rate. Conversely, the incidence rate or risk of developing hepatitis could be greater in B yet its prevalence less than that in A, because the people who get hepatitis in B die quicker than those in A.

b. Benefits and Disadvantages of Case-Control Studies

One advantage of a case-control study is that rare diseases can be studied without using large sample sizes. The reason for this, as discussed later, involves the statistical tests that are performed to determine whether the study is statistically significant. Therefore, a case-control study can be considerably less expensive than a cohort study when evaluating rare diseases.

The case-control study also has an advantage over a cohort study with respect to studying latency. Assume that a plaintiff alleges that exposure to plastic caused her to develop heart disease. Assume that a cohort epidemiological study conducted on people who had been exposed to plastics for an average of 10 years concluded that there was no increased risk of developing heart disease from exposure to plastics. The plaintiff's experts will argue that the study is either uninformative or does not address the plaintiff's allegation because (in the opinion of the testifying expert) the latency period needed to develop heart disease following exposure to plastics is 20 years. The expert argues that since the study lasted only 10 years, it would not have revealed cases of heart disease that develop over 20 years.

In contrast, a case-control study would have begun with patients who had already developed heart disease. Therefore, whether the latency is 1, 10 or 20 years is irrelevant for a case-control study because

15. *See* Gordis, *supra* note 4, at 32, 36.

whatever latency period is necessary has already occurred when the study begins with people who have developed the disease. Therefore, a case-control study takes the latency argument out of the arsenal of the plaintiff's expert.

Among the disadvantages of a case-control study is that it cannot determine the incidence rate of disease. The reason is that the incidence rate in the exposed population is not known since the study design begins with people who already have the disease.[16]

The second disadvantage of case-control studies is that they are more susceptible to selection bias than a cohort study. For the data from a case-control study to be "generalizable" (*e.g.*, applicable to the general population), the cases must fairly resemble people who develop the disease in the general population. In other words, the cases must be homogenous with the general population who develop the disease. If the disease severity in the test group is different in some respect from people who develop the disease in the general population, then the study may yield inaccurate data. Second, there must be some uniformity in the diagnostic criteria utilized to select cases for the study. Stated differently, rigid diagnostic criteria must be utilized to ensure that all of the cases meet common diagnostic criteria.

Disadvantages of a case-control study also exist with selecting controls. Some authors believe that selecting controls in case-control study is perhaps the most critical feature of the study.[17] Controls should be similar in all important respects to the cases except for disease status. Even if the controls are properly matched to the cases, for the data to be generalizable, the controls must be representative of the general population. If the controls, although properly matched to the cases, are nevertheless different in an important respect from the general population, the data may not be reliable.[18] For example, controls are often selected from hospital populations. Depending on the exposure and diseases being studied, such a population may lead to a bias either in favor of falsely increasing the risk or masking a real association.

For example, suppose a case-control study is designed to study the association between coffee and stomach cancer. A control population is drawn from hospital patients who do not have stomach cancer but who are hospitalized for other ailments, such as high blood pressure. This control population may have reduced their coffee consumption based on physicians' orders having to do with their high blood pressure, and therefore their exposure may be artificially low relative to the general population. Such a study would underestimate coffee consumption in the general population and therefore would misleadingly increase the risk between coffee and stomach cancer.[19]

3. Nested Case-Control Studies

Because there are benefits to both cohort and case-control studies, it is only logical that a study combining the benefits of both provides potential advantages over either. Such is the case with a nested case-control study, which is a case-control study that begins as a cohort study with neither exposed nor non-exposed groups having developed the disease. Over time, some number of the exposed and control groups may develop the disease of interest. Those in both groups who develop the disease (a "nested" group) are regrouped as the case population. A sample from the original cohort that had not developed the disease is regrouped as controls. The percentage of the cases that had been previously exposed is calculated and contrasted to the percentage of controls who had been exposed. A comparison of the rates will reveal if there

16. *Id.* at 145.
17. *See* Hennekens & Buring, *supra* note 7, at 137.
18. *Id.* at 137-38.
19. *Id.* at 138-39.

was an increased risk among the cases from having been exposed.[20]

C. Factors Creating Unreliability

There are three factors that can affect the results of an epidemiological study, making the data unreliable. When reviewing studies or examining experts who have relied on epidemiological data, counsel need to examine the role, if any, that bias, confounding factors or random chance may have had in producing the results.

1. Bias

Bias can be broadly defined as any effect during a study that produces erroneous data. For example, for many conditions, age can be an important variable in determining the susceptibility to a disease or exposure of interest. By not controlling (*e.g.*, selecting appropriate cases and controls) for this variable, the study may over- or underestimate the true risk of the exposure in producing the disease of interest.

When counsel are interpreting an epidemiological study, careful attention must be given to whether the study has been subject to bias of any type. Both clinical and observational epidemiological studies are susceptible to bias. Bias can be minimized through appropriate steps taken during the design and analytical phases of a study. Plaintiffs' experts who rely on an epidemiological study should be questioned about bias, and defense experts should be prepared to defend a study from attacks of bias. The effect of bias cannot be quantified. Therefore, unlike the role of chance or confounding factors, whose effects can be quantified, the effects of bias cannot be negated by a statistical analysis.

When deposing plaintiffs' experts or preparing defense experts for deposition or trial, counsel should review with the experts all forms of bias that may have potentially affected a study. Although many forms of bias have been identified,[21] the following are the major forms of bias in cohort and case-control studies.

a. Selection Bias

Selection bias is defined as a process through which "study subjects are selected in a way that can misleadingly increase or decrease the magnitude of an association."[22] As discussed earlier, cases and controls must be selected in the same manner. Consequently, if investigators select cases in a manner utilizing different criteria than used to select controls, then the study is subject to selection bias, and the data may not be valid. If in a historical cohort study, investigators select the exposed subjects knowing that they had already developed the disease being studied, then selection bias has occurred, and the data probably are misleading.[23]

Another form of selection bias, often referred to as response bias, may occur when subjects are recruited by a questionnaire. The degree of willingness of someone to participate in a study may reflect a special interest that makes the volunteer unlike the general population. For example, individuals responding to a questionnaire who are being recruited as controls may be reflective of a group whose attention and sense of responsibility might reflect seeking better medical attention resulting in an improved state of health than others in the general population. A control that is healthier than the general population would tend to increase the association between exposure and the test group, because fewer members of the control group will end up developing a disease.

Questionnaires sent to different socioeconomic neighborhoods will produce controls that may affect the outcome of the study. It is reasonable to assume that wealthier people have access to better

---

20. For a full discussion of the benefits and disadvantages of nested case-control studies, *see generally* Gordis, *supra* note 4, at 136-37.

21. *See* David L. Sackett, *Bias in Analytical Research*, 32 J. CHRON. DIS. 51 (1979).

22. GARY D. FRIEDMAN, PRIMER OF EPIDEMIOLOGY 173 (3d ed. 1987).

23. Hennekens & Buring, *supra* note 7, at 273.

medical care and therefore may be healthier or more aware of their symptoms and exposures. Such a population may or may not reflect the general population.

Selection bias also may occur when the selection process is conducted by interviews. Interviewers aware of the study hypothesis may seek subconsciously to enroll cases or controls that, depending on the study hypothesis, are either more or less likely to have been exposed.

Even if the study populations are different among themselves or the general population in ways other than the variables being studied, that does not necessarily mean that bias invalidates the epidemiological data. Before an expert can conclude that the study data is invalid, an assessment must be made as to whether the bias had any effect, and, if so, whether the effect was more likely to increase or decrease the exposure risk. For example, a study might very well be subject to selection bias. However, if the bias would be expected to increase an association between exposure and disease, then the existence of bias is irrelevant in a study that found no increased risk of exposure and disease.[24]

b. Information or Measurement Bias

Information or measurement bias results when the method of data collection "makes the information obtained from two or more groups differ in some misleading way."[25] Bias can occur in different ways during the process of collecting information. For example, in a case-control study, the investigators must determine who among the cases and controls have been exposed. People with a disease often have a better recollection of exposures that they have been told or read may have caused their disease. Not surprisingly, healthy controls are less likely to recall exposures that had no significance to them when they occurred.[26]

Information bias also may occur as a result of misclassification resulting from improper recording of medical diagnosis in medical records and death certificates. Death certificates are notoriously unreliable in recording the true causes of death. By law, death certificates state the underlying "cause" of death, which is defined as that which "initiated the train of morbid events leading directly or indirectly to death or the circumstances of the accident or violence which produced the fatal injury."[27] Death certificates often do not record the immediate cause of death. In either case, relying on death certificates to obtain epidemiological data can produce unreliable results.[28]

Misclassification of employment records also may be a source of bias. For example, exposure may be reasonably inferred from certain occupations. If the occupational status of controls is recorded inaccurately, their exposure will be underestimated. Such a mistake could erroneously increase the risk assessment. Therefore, information bias may affect both the incidence of disease and exposure history.

All studies have some degree of misclassification. When the misclassification is random, the bias probably will not affect the data because the bias affects both cases and controls. Thus, if a plaintiff's expert criticizes a study because of measurement bias, counsel should question the expert on the basis for his or her presumed conclusion that the bias disproportionately affected only the cases or controls.

Another form of information bias is interviewer bias. Randomized clinical trials that are double blind are not subject to interviewer bias because of the double blind nature of the study. All forms of observational epidemiological studies, however, are potentially subject to interviewer bias. Interviewer bias occurs when there is a "systematic difference in the soliciting, re-

---

24. *See id.* at 282-84 for a further discussion on evaluating the role of bias.
25. Friedman, *supra* note 22, at 273.
26. Hennekens & Buring, *supra* note 7, at 274.
27. *See* Gordis, *supra* note 4, at 46.
28. *See also* Friedman, *supra* note 22, at 34.

cording or interpreting of information from study participants."[29]

For example, in a case-control study in which the interviewer is aware that the subject is a case and not a control, such knowledge may affect the interviewer's questions. Depending on the study hypothesis, the interviewer may not be equally aggressive seeking exposure data from cases and controls. Depending on the questioning, interviewer bias may increase or decrease the true relative risk of the exposure.[30]

Exploring the many forms of bias is a fruitful area of inquiry when deposing plaintiffs' experts. The occasion arises when those experts are relying on epidemiological evidence to support an opinion on causation, as well as when experts are attacking epidemiological studies that have failed to find an alleged association between exposure and disease. In both instances, defense counsel must have a good understanding of bias and the circumstances under which it can affect the validity of epidemiological data.

2. Confounding Factors

Non-infectious ("chronic") diseases are rarely the result of one causative agent. Their etiology is more accurately viewed as a "web of causation," with many dependent and independent factors responsible for the onset of disease.[31] Because epidemiological studies are generally designed to examine only one suspected risk factor, the study must be carefully evaluated to determine whether it is design controlled for variables related to the exposure and disease. Such variables are called "confounding factors," because they may falsely increase the risk or mask the true risk between exposure and disease.

A confounding factor is a variable that is independently related to the exposure and disease. For example, assume a study is done to determine the relationship between exercise and heart disease. Because age is related to the level of exercise one gets and independently related to heart disease, a study looking only at the relationship between exercise and heart disease will produce erroneous results unless the study has controlled for the confounding factor of age. A common sense example offered in the Federal Judicial Center's *Reference Manual on Scientific Evidence* is the relationship between infant shoe size and vocabulary. One would expect to find, if this hypothesis were tested, that there is a correlation between increasing shoe size and improved vocabulary. However, the confounding variable of age would explain the correlation better than the increased shoe size.[32]

Counsel should examine a study closely to see if the investigators controlled for confounders when the study was designed. For example, if age is a variable, then look to see if controls were selected that matched the cases by age. A study that controls for confounding variables contains better epidemiological data.

Another approach, although less preferable, is to "control" for confounders by quantifying their effect on the data by using appropriate statistical tests, such as stratification analysis and multivariate analysis.[33] Counsel should look to see if such analysis was done if it appears the design did not control for confounders. The better the data, the more accurate the results. Therefore, data produced by a well-designed study that has controlled for confounders is better than data subject to confounders whose effect must be "removed" by statistical tools.

3. Random Chance

Random chance is a third factor that can systematically affect the results of a study in a way that falsely increases or decreases the association between exposure and dis-

29. Hennekens & Buring, *supra* note 7, at 275.
30. Fletcher, *supra* note 3, at 275.
31. Friedman, *supra* note 22, at 3.
32. Reference Manual, supra note 8, at 365.
33. Hennekens & Buring, *supra* note 7, at 304-21.

ease. If everyone in the general population could be sampled, then random chance would not explain the results derived from a study. However, observational epidemiological studies are conducted by sampling a small group of the general population, and then on the assumption that the cases and controls are representative of the general population, the results are extrapolated to the general population. As a result, in all epidemiological studies, chance exists as one of the explanations for the results of an epidemiological study.

It is believed that most chronic diseases have more than one cause. Moreover, most of the causes and the synergistic effects between various causes, such as genetics and the environment, are unknown. Therefore, scientists believe that virtually all diseases occur in the non-exposed general population at a baseline rate. Because the "web of causation" is unknown for the great majority of diseases, there is no absolute certainty that a subject who has been exposed and who subsequently develops the disease endpoint did so because of the exposure or the presence of other unknown etiologic factors.

Stated differently, in any individual case, one does not know whether the individual would have developed the disease by chance regardless of the exposure. To determine the probability that chance produced the results, epidemiologists use a variety of statistical tests to evaluate the data. The appropriate test is a function of study design, the nature of the data and the distribution of the data.

Data points derived from the study can be plotted on different types of graphs to determine their distribution. Data that are "normally" distributed are symmetrical and appear as a bell-shape curve. The variability or spread of data that are normally distributed can be described by calculating the standard deviation. Depending on the data distribution, tests are selected to compute the *P* (probability) value.

Clinical data is generally not normally distributed.[34] Not uncommonly, physicians use a Students t-Test to evaluate biological data despite the fact that such a test should be used only on normally distributed data. Therefore, counsel should look to see whether the data are normally distributed and whether the test used was appropriate.

Finally, data may be continuous or discrete. Continuous variables are those without discrete values. If one asked how many people in high school drank Coke versus Pepsi, the answer would produce discrete data. Conversely, data that are characterized as "low, medium and high" are continuous. A Students t-Test is used for continuous data and a Chi-Square test for discrete data.[35]

Traditionally, the role of chance in producing the observed results has been evaluated by calculating the *P* (probability) value. As discussed later, the *P* value is defined as the calculated probability of getting by chance alone the results obtained, assuming there was no association between exposure and disease.[36] Many epidemiologists, however, have begun to rely more on the confidence interval to interpret the significance of data. Although a subject of debate, many epidemiologists believe confidence intervals provide a more quantitative assessment of the association between exposure and disease. Further on, this article discusses *P* values and confidence intervals.

WHAT DOES ALL THIS MEAN?

Assume that, after carefully reviewing a study, defense counsel concludes that the study was free of bias and confounding factors. Counsel is presented with an array of mind-numbing statistics from which one is to draw sound conclusions with respect to whether chance might be an appropriate explanation for the purported association. The following should help one to wade through the maze of statistical data to understand the statistical jargon often found

34. *See* Fletcher, supra note 3, at 33.
35. *See* Reference Manual, *supra* note 8, at 379.
36. *Id.* at 402.

in epidemiological studies. Included is a checklist of questions that counsel should consider in preparing for a deposition of the plaintiffs' epidemiologists or when consulting with defense epidemiologists.

A. How Are Data Presented?

1. Mean, Median and Mode

Epidemiological data are obtained by sampling a population thought to be representative of a larger population. A question counsel should ask is how the data have been presented and whether the manner fairly represents the true values.

A set of numerical values (*e.g.*, data points) can be presented in a number of ways. The most common is to calculate the "mean" of the values. The mean is the arithmetic average of the numbers. In cases involving large sets of numbers or small sets that do not contain "outliers" (unusually large or small values), the mean is a good way to present the data. However, the mean of a group is particularly sensitive to large or small extreme values. Such values will disproportionately move the mean up or down relative to the average of the majority of the values.

A more accurate way to present the data may be to use the median value, which is that number in which half the numbers in the sample are greater and the other half are less. Using the median value eliminates the disproportional effect of outliers in a small sample. However, the median value does not give the reader any information about the values above and below the median, and it is a less effective way of calculating the statistical significance of data.[37]

Finally, the mode is that number most commonly found in the sample.[38] When reviewing an epidemiological study, consider whether using a different value than that used by the authors would produce a more accurate representation of the spread of the data.

2. Standard Deviation

When a sample is taken of a larger population, resulting values (which produce mean, median and mode values) will vary slightly with each successive sampling of the larger population. With normally distributed data, the standard deviation describes the spread of individual data points.

Although often confused with the standard deviation, the standard error measures the variability in the mean of the values in each sample. The standard error of the mean permits inferences to be made about the larger population, because it reflects the expected variability in the mean that would result with successive sampling of the larger population. Stated differently, the standard error of the mean reflects, "the degree to which the sample mean reflects the true population parameter."[39]

Standard deviation describes the variability of individual data points around the mean of the data. One, two and three standard deviations refer to the area or region in which 68, 95 and 99 percent, respectively, of the individual data points would fall.

3. Relative Risk (RR) and Odds Ratio (OR) and Attributable Proportion of Risk (APR)

An epidemiological study examines the statistical strength of an association between exposure and disease. An exposure is a risk factor if it is associated with the disease to a degree of statistical significance.

In a cohort study, a percentage of the exposed and non-exposed patients will develop the disease being studied. The percentage of exposed subjects who develop disease divided by the number of exposed subjects who do not develop the disease produces a risk ratio of disease among the

37. *See generally* Hennekens & Buring, *supra* note 7, at 233.

38. *See generally* Reference Manual, *supra* note 8, at 360.

39. *See* Hennekens & Buring, *supra* note 7, at 239-40.

exposed population. Similar risk ratios are calculated for the control population that develops the disease. To determine if the exposure produced an increase in the risk ratio of disease as a result of exposure, the risk ratio of the exposed and non-exposed populations are compared. The resulting quotient is called the relative risk (RR). The relative risk is obtained by dividing the incidence of disease in exposed by the of disease in unexposed.

In other words, the RR is a measure of the degree to which it is more likely that an exposed population will develop disease than those who are not exposed.[40] For example, assume a study enrolls 100 coffee drinkers and 100 non-coffee drinkers. If 50 of the exposed and 20 controls develop high blood pressure then the RR is calculated as follows:

	High Blood Pressure	No High Blood Pressure	
Coffee Drinkers	A	B	n =
	(50)	(50)	100
Non-Drinkers	C	D	n =
	(20)	(80)	100

$$RR = \frac{A/(A+B)}{C/(C+D)} = \frac{50/100}{20/100} = 2.5$$

From this data, coffee drinkers have a 2.5 higher risk of developing high blood pressure than non-drinkers.

An RR of less than one means that the exposure (assuming a causal relationship) protects against the disease developing. If RR = 1, then the exposure has not increased the risk of disease.[41] Conversely, if, as in the example above, RR > 1 (assuming that the increase is statistically significant), then the exposure is associated with an increase in the risk of the disease.

Although the RR quantifies the increased risk of developing a disease endpoint with exposure, it does not show how many new cases can be expected. While the RR of an exposed population may be large, the absolute risk for the exposed population developing the disease may nevertheless be extremely small for diseases which are very rare. Plaintiffs' experts in the above example will argue that as a result of exposure, "two and a half times" more cases will develop than among those who have not been exposed. Although this is an accurate statement of the RR, defense counsel may want to put the statement in context in terms of the actual number of disease cases that would be expected assuming there was a causal relationship between exposure and disease.

Such data is derived by calculating the attributable proportion of risk (APR). The APR calculates the number of new cases that are expected over the number expected in the non-exposed population (*e.g.*, background rate). The APR is calculated by subtracting the rate of disease in the non-exposed group from the rate of disease in the exposed group. For example, if the rate of high blood pressure in males 30-50 who drink two or more cups of coffee a day is 10 per year per 100 patients, and if the rate of high blood pressure in non-coffee drinking males ages 30-50 is 8 per 100 per year, then two additional cases of high blood pressure per year per 100 men ages 30-50 can be attributed to coffee drinking.

Although the APR can provide valuable public health information, it does not address the strength of the association between exposure and disease. In the previous example, although two additional cases of high blood pressure may result

40. Fletcher, *supra* note 3, at 105; Hennekens & Buring, *supra* note 7, at 77.

41. This article discusses later the court decisions that have excluded statistical evidence of a relative risk of less than 2, since such evidence does not satisfy the legal standard of preponderance of the evidence. The RR measures the observed association between exposure and disease; it does not address deficiencies in the evidence resulting from bias and confounding variables. Therefore, when interpreting an epidemiological study, one must always evaluate the quality of the data for bias and confounding variables before considering the RR as a reliable indicator of risk.

from drinking coffee, one would not know anything about the likelihood of developing high blood pressure from drinking coffee. That information is provided by the RR.

Because a case-control study begins with a group that has developed the disease, it is not possible to determine the rate at which the disease developed in the group. Stated differently, there is no risk that the subjects will develop the disease, because they already have the disease. Therefore, a RR cannot be computed from the data generated by a case-control study. Instead, risk assessment in a case-controlled study is calculated by the odds ratio (OR).

Although the terms "odds" and "probability" are often used interchangeably, they mean different things. The probability that an event will occur is not the same as the odds that the event will occur. For example, if the probability of something happening is 80 percent, the probability of the event not happening is 20 percent. The odds of the event occurring are defined as "the ratio of the number of ways the event can occur to the number of ways that the event cannot occur".[42] To calculate the odds that the event will occur based on a probability of it occurring at 80 percent is calculated by dividing the probability that it will occur by the probability that it will not occur.

Using the previous hypothetical for RR, the OR is calculated as follows:

	High Blood Pressure	No High Blood Pressure	
Coffee Drinkers	A (50)	B (50)	n = 100
Non-Coffee Drinkers	C (20)	D (80)	n = 100

$$OR = \frac{AD}{BC} = \frac{(50)\ (80)}{(20)\ (50)} = 4$$

In most epidemiological studies, the OR is a good estimate of the RR if the following three conditions are met: (1) the cases are representative of all cases in the general population, (2) controls are representative with respect to exposure to the general population, and (3) the rate of the disease is low.[43]

4. Alpha and Beta Errors

The first step in determining the probability that chance may have produced the association between exposure and disease as reported by an epidemiological study begins with understanding Alpha and Beta.

a. Alpha

Alpha is the level of statistical probability that is arbitrarily chosen by the investigator before the study begins. If the data satisfies or meets the Alpha level, then the results are said to be statistically significant.[44] Customarily, scientists select an Alpha level of .05. If the level of .05 is reached by one of the statistical tests discussed, it means that there is only a 1 in 20 probability that the results were obtained by chance alone assuming there is *no* association between exposure and disease. Scientists generally are not willing to conclude that there is a causal link between two events that have more than 1 in 20 probability of occurring by chance alone.

When reviewing an epidemiological study, careful attention must be given to

42. Gordis, *supra* note 4, at 145.

43. In the example above, high blood pressure had a high incidence rate, which creates a disproportionate relationship with the odds ratio. *See generally* Gordis, *supra* note 4, at 147-48; Hennekens & Buring, *supra* note 7, at 80-81. Some authors assert that a disease must have an incidence rate of greater than 1/100 for the distortion between RR and OR to become significant.

44. Fletcher, *supra* note 3, at 218.

where Alpha has been set, because by definition an investigator can claim that the results are statistically significant provided they meet the level of Alpha. If, however, an investigator has chosen an Alpha level higher than the customary .05, then a claim of statistical significance must be reviewed in the context of a study that has departed from conventional and accepted scientific standards.

If data meet the Alpha level (*i.e.*, statistically significant), the investigator may conclude that because the probability is so low that the association observed was the result of chance (*e.g.*, the probability of this happening is less than 1 in 20), it must be real. In epidemiological jargon, the "null hypothesis" of no association has been "rejected." In other words, if among a population of coffee drinkers there is a statistically significant increase in blood pressure compared to controls, the investigators can reject the null hypothesis and conclude that the study demonstrates an association between coffee drinking and high blood pressure.

It is tempting, but incorrect, to conclude that if a study has met an Alpha level of .05, there is only a 5 percent chance that the null hypothesis (no association between drinking coffee and high blood pressure) could be correct. The hypothesis of whether coffee drinking causes high blood pressure is either true or false. Concluding that data are statistically significant only describes the probability of the events occurring by chance. The conclusion that data are statistically significant provides only a guide to determining whether the hypothesis is true or false.[45]

Because the entire population is never tested, even when data is statistically significant at an Alpha of .05, chance can never be totally eliminated as an explanation for the observed association. For example, in the preceding example, although the rate of high blood pressure among coffee drinkers exceeded the rate among nondrinkers when Alpha is .05, there still remains a 1 in 20 chance that the association was due to chance. If in fact the association is due to chance, then concluding that an association exists is a "false positive" result. A false positive association is called an Alpha or Type I error.

b. Beta

Beta is a measurement of the probability of falsely concluding that there is no difference between exposure and disease, when in fact a true association exists. A Beta error (false negative or Type II) is one in which investigators "accept" the null hypothesis of no association, when in fact an association exists. Generally, for most epidemiological studies, Beta ranges between 10-20 percent. The considerably lower level of Alpha reflects the preference of scientists to minimize the risk of falsely concluding that an association exists when none really exists.

Whereas Alpha can be arbitrarily set by an investigator, Beta is a function of a mathematical calculation that depends in part on the Alpha level. As such, there is an inverse mathematical relationship between Alpha and Beta. Consequently, reducing Alpha (thereby reducing the probability of a false positive association) will increase the probability of a false negative (Beta) error.

5. *P* Value and Confidence Intervals

a. *P* Value

The *P* (probability) value is derived by analyzing data with an appropriate statistical test. If the *P* value is equal to or less than the pre-set Alpha level, then the data are said to be statistically significant.[46] In many epidemiological studies, the term "*P* value" is used interchangeably with Alpha. The statement that, "the data were statisti-

45. *See, e.g.*, Reference Manual, *supra* note 8, at 379, 381; Hennekens & Buring, *supra* note 7, at 33.

46. *See* Reference Manual, *supra* note 8, at 402; Hennekens & Buring, *supra* note 7, at 244-45.

cally significant at a *P* value of .05," technically means that the data were analyzed at an Alpha level of .05 and that the results of this statistical analysis yielded a *P* value that was at or below .05.

A number of statistical tests, such as the Student's t-Test, Chi-Square and Fisher's Exact Test, are used to calculate the *P* value. As discussed earlier, the distribution of the data, among other factors, will affect the statistical test to be used. For example, a Student's t-Test can be used only for data that are continuous and Gaussian (*e.g.*, normally distributed).[47] Not infrequently, an expert may rely on data not normally distributed and that the authors have analyzed with an inappropriate statistical test. In some circumstances, re-evaluation of the data with the appropriate test may yield different results.

Finally, most statistical tests can be said to be either "1-tail" or "2-tail." For reasons having to do with statistical computations, it is easier for data to achieve statistical significance (regardless of the Alpha level), if the test is 1-tail.[48] As a practical matter, most statistical tests done on biological data are 2-tail. An epidemiological study, using a 1-tail test, that has failed to show an increased RR is a more powerful statistical statement than if the analysis was done using the more customary 2-tail test.

b. Confidence Interval

Since the *P* value is only a statement of the probability of an event occurring in a sample of a much larger population, some argue that relying exclusively on the *P* value gives either more or less value to epidemiologic data than is warranted. In part, the *P* value is a function of the size of the sample being studied. Therefore, a large difference can exist that is not statistically significant because of the small sample size. In this situation, it would be misleading to conclude there was no difference.

Assume, for instance, that after five at-bats in a new season, two baseball players are hitting .400 and .275, respectively. One could not conclude the .400 batter is a better hitter despite the large difference, because the sample size (at bats) is too small. The difference would be said not to be statistically significant. Conversely, if after 200 at-bats, the two players' averages are .280 and .275, one can draw a conclusion. Because the sample size is large, the difference between .280 and .275 may be statistically significant.

However, few baseball fans would say the difference of .005, although statistically significant, is a meaningful basis on which to conclude that the .280 player is a better hitter. Similarly, determining whether epidemiological data that are statistically significant have any biological significance requires an analysis of other factors.

Because of the limitations of relying solely on the *P* value, many studies now use a confidence interval to describe data. A confidence interval is expressed as the mean value ("point estimate") plus or minus the upper and lower limits of the confidence interval. For example, if the average age of players on a professional football team is 25, with 95 percent of the ages falling between 23 and 27, the confidence interval is 25 +/- 2.[49]

A confidence interval also conveys whether or not the range is statistically significant. Because an RR of 1 means that there is no increased risk, a confidence interval whose lower limits include one within the range of the interval by definition means that the range is not statistically

---

47. *See* Hennekens & Buring, *supra* note 7, at 246-49; Fletcher, *supra* note 3, at 33, 190-92.

48. *See* Reference Manual, *supra* note 8, at 382.

49. The confidence interval need not be symmetrical. A confidence interval can be expressed as RR = 2.5 with a confidence interval of (1.4-6.5). If the Alpha was set at .05, this would mean that the best estimate of the RR is 2.5, but that with 95 percent confidence, the risk is not lower than 1.4 and no greater than 6.5.

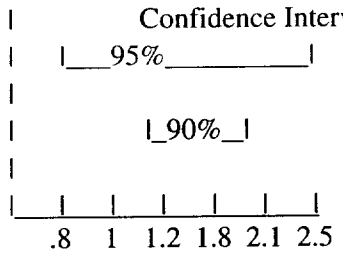
significant. For example, a confidence interval of 1.5 +/- .6 means that the RR will fall in the range .9 to 2.1.

Because the lower boundary of the confidence interval falls below 1, the result cannot be statistically significant, since the exposure could be either protective or a risk factor. Conversely, if the confidence interval was 1.5 +/- .4, then the result is statistically significant since the lower boundary of confidence (1.1) exceeds 1.[50]

Finally, because Alpha is set arbitrarily, an investigator is free to set it at levels above .05 (*e.g.*, confidence levels below 95 percent). Increasing the Alpha level from .05 to .10—or, stated differently, reducing the confidence level from 95 to 90 percent—means that one is less sure that the true RR falls within the range. As a result, the size of the interval decreases.

For example, assume a RR is computed to be 1.8, which, at a 95 percent confidence level, has a range of .8 to 2.5. If, as shown in the diagram below, the confidence level is reduced to 90 percent, the confidence interval also will get smaller because there is less confidence that the true RR falls in the interval. Plaintiffs' experts frequently will re-analyze an epidemiological study whose RR is not statistically significant and will recalculate the confidence interval after increasing Alpha. Alpha is increased to the point needed to increase the lower boundary of the confidence interval above 1. At that point, the experts claim that study results are statistically significant. The expert will proclaim that the statistically significantly study demonstrates that the exposure caused the disease which the plaintiff has suffered.

This approach should be challenged under *Daubert* as inconsistent with the methodology typically and customarily followed, which establishes Alpha at .05. Stated differently, as the level of predicted certainty with the RR increases, the margin of the point spread also increases.



6. Power

Power is the statistical strength of a study to detect a weak association. Power establishes the Beta level of a study.[51] The power of a study is directly correlated to disease prevalence and sample size. Therefore, the power of a study increases with an increase in sample size and/or prevalence.

Power calculations are relevant during the design phase of a study. If one knows the disease prevalence and the cohort size, one can calculate the smallest RR the study should be capable of detecting, assuming an association exists, for the sample size. For example, the statement that a study has a 90 percent power to reveal a minimal detectable RR of 3 means that there is a 90 percent probability that the study could detect a RR as low as 3.[52] If such a study was performed, it could not be expected to rule out, at a level of statistical significance, a RR between 1 and 3.

Power is a useful mechanism by which to plan epidemiological studies to ensure that the size of the study is sufficiently large to provide meaningful data on the as-

50. Hennekens & Buring, *supra* note 7, at 252-53; Fletcher, *supra* note 3, at 198.

51. *Id.* at 259. Power = 1-Beta.

52. Steven N. Goodman & Jesse A. Berlin, *The Use of Predicted Confidence Intervals When Planning Experiments and the Misuse of Power When Interpreting Results*, ANNALS INT. MED. 200, 201 (8/94).

sociation (if one exists) between exposure and disease.[53]

As discussed later, plaintiffs' experts often perform *post hoc* power calculations on "negative" epidemiological studies in order to argue that the study should not be interpreted to mean there is no association because the study had insufficient power to detect such an association. However, once a study has been completed and a RR point estimate and confidence interval established, the power of the study is irrelevant. The confidence interval provides the data of what was actually shown to be the association in the cases and controls.[54]

7. Multiple Comparisons

It is often not revealed in epidemiological studies how many comparisons of exposure and different disease endpoints were performed before a reported association was observed. The process of pouring over exposure data with countless different disease endpoints is referred to as "data dredging." As more comparisons are conducted, the probability of finding one comparison that is statistically significant increases.[55] Therefore, when examining an expert who proffers an opinion on causation based on epidemiological data, one area of inquiry should be the number of comparisons and the type of comparisons that were conducted before the putative association was identified. If associations are revealed after multiple comparisons, there are appropriate statistical tests that can be done to correct for the effect of multiple comparisons.

Counsel should always question witnesses who are relying on epidemiological evidence of an association between exposure and disease to determine if the appropriate statistical corrections were made. In general, the probability of a false positive (Alpha) error increases as the number of comparisons are increased. Conversely, the probability of a false negative error increases as the statistical power of a study decreases.

8. Role of Epidemiological Data in Assessing Causation

Assume that an epidemiological study has shown a RR of 1.5 between coffee drinking and high blood pressure with 95 percent confidence intervals of 1.1-1.8. A plaintiff's expert will argue that this statistically significant data demonstrate that coffee drinking has a large impact on high blood pressure, since it increases the risk by 50 percent (1 to 1.5). The expert concludes that based on such "overwhelming evidence," there is no question in his mind that the plaintiff's coffee drinking caused his high blood pressure.

This reasoning is fallacious for a number of reasons. First, a RR of 1 means that the exposure has not accounted for any more cases of the disease in the exposed population than that which occurred in the non-exposed population. A RR of 2 means that the exposure has produced an equal number of cases as the background rate. Therefore, to prove by the preponderance of the evidence (*e.g.*, more than 50 percent probability) that exposure caused the disease, a plaintiff must prove a RR of something more than 2.[56] In the example above, it is more likely than not that reasons unrelated to the risk factor, assuming there is a causal relationship, caused the disease. Even if, however, the RR of an epidemiological study exceeds 2, such evidence alone does not prove causation.

Epidemiological evidence can demonstrate only a statistical association between

53. *Id.* at 200; Alan H. Smith & Michael N. Bates, *Confidence Limit Analysis Should Replace Power Calculations in the Interpretation of Epidemiologic Studies,* EPIDEMIOLOGY 449 (8/92).

54. *Id.* at 449; Goodman, *supra* note 52, at 201-02; Fletcher, *supra* note 3, at 200.

55. Fletcher, *supra* note 3, at 200-02.

56. *See Daubert v. Merrell Dow Pharmaceuticals Inc.*, 43 F.3d 1311, 1320 (9th Cir.), *cert. denied*, 116 S.Ct. 189 (1995) (relative risk of less than two tends to disprove legal causation, since it does not double likelihood that Bendectin causes birth defects). *See also* Reference Manual, *supra* note 8, at 169 n.128.

exposure and disease. Whether the association is causal depends on an analysis of additional factors. In 1969, Bradford-Hill proposed a number of factors to consider in assessing the causal relationship between exposure and disease.[57] These factors, as modified over the years, are:

1. Strength of the association. (Is there a significant RR for the exposure?)
2. Temporal relationship. (Has the exposure preceded the disease?)
3. Consistency of the association. (Have the epidemiological studies or other biologic data been consistent in their conclusions with respect to the possible or causal role of the exposure to the disease? Have different labs generated consistent data using different methodologies?)
4. Biological plausibility. (Is the proposed association consistent with biological understanding?)
5. Alternate explanations. (Have potential biases and confounding factors been examined and ruled out as well as other biological explanations?)
6. Dose response.[58]
7. Specificity. (Has the agent been associated with a single disease or type of disease?)[59]

Finally, different types of causal relationships have been proposed which may have significance depending on the law of comparative negligence and joint and several liability in a particular jurisdiction.

1. Necessary and sufficient (analogous to the "but for" test of causation).
2. Necessary but not sufficient (disease will not occur unless there was exposure, however, exposure alone is insufficient).
3. Sufficient but not necessary (exposure alone can produce disease; however, disease may also result from exposure to other agents).
4. Neither necessary nor sufficient (the disease may occur in the presence of exposure but may also occur in the absence of exposure).[60]

QUESTIONS TO CONSIDER IN INTERPRETING AN EPIDEMIOLOGICAL STUDY

When a statistical association between exposure and disease is reported by an epidemiological study, a number of questions should be considered. Among these are:

1. Was the study well-designed?

In a case-control study,

- Were the cases selected representative of the general population?
- Was the severity of the diseases experienced by the cases representative of the disease severity in the general population?
- Were uniform diagnostic criteria utilized to identify the cases?
- Were the cases selected with knowledge of their exposure status?
- Did the study include both old (prevalence) and new (incidence) cases?
- Were the controls selected representative of the general population?
- Were the controls selected from a hospital or general population? If selected

57. A.B. Bradford-Hill, *The Environment and Disease: Association or Causation?* 58 Proc. R. Soc. S. Med. M. 295-300 (1965); Fletcher, *supra* note 3, at 241.

58. It is commonly accepted in medicine that for most etiologic agents, a dose response relationship exists. However, some argue that dose response is of lesser value in assessing a causal relationship since the threshold level necessary to induce disease may be beyond the levels of detection in biological assays or the sensitivity of epidemiological testing. *See* Hennekens & Buring, *supra* note 7, at 43.

59. Many argue that this factor is of least importance in assessing causation because of the multifactorial aspect of chronic diseases. A much stronger argument can be made for causation if specificity exists. For discussions on Bradford-Hill's criteria, *see generally* Hennekens & Buring, *supra* note 7, at 39-43; Reference Manual, *supra* note 8, at 160-66; Fletcher, *supra* note 3, at 241-45.

60. Reference Manual, *supra* note 8, at 164-65.

from a hospital setting, was their illness a confounding variable?

- Were the controls selected with knowledge of their exposure status?
- Was the exposure measured in a valid and reproducible manner?

In a cohort study,

- Was the population selected sufficiently similar to the general population so that the data is generalizable?
- Have disproportionate numbers from either the exposed or non-exposed been lost during the course of the study?
- Does the study length correspond to the latency of the disease studied?
- Was the selection of the exposed and non-exposed subject to selection bias?

For either a case-control or cohort study,

- Was the information from which exposure was obtained reliable?
- Was the study subject to media bias, litigation bias or interviewer bias?
- Were all confounding factors identified and controlled for in the design? If confounding factors were not controlled for in the design, has their effect been evaluated by appropriate statistical analysis?

2. Could the observed results be the result of a chance?

- Was the Alpha level set at the customary level of .05 or at some higher level?
- Although not statistically significant, did the difference in the two populations nevertheless appear to be real given the point estimate and the width of the confidence interval?
- Does the data appear to be normally distributed and if not, were the appropriate statistical tests done to either normalize the data or to otherwise evaluate it?
- If the data is claimed not to be statistically significant at a *P* value of .05, would the difference otherwise be important if a confidence interval had been calculated?
- Was the presentation of the data as a mean, median or mode misleading?
- Could the statistical average as represented by either the mean, median or mode been affected by outliers?
- Was a claim of statistical significance (or lack thereof) achieved by using a 1-tail or 2-tail statistical test?
- Was a claim of statistical significance the result of multiple comparisons? If so, were appropriate statistical corrections performed on the data?
- Is the claim of a statistical significant relationship nevertheless biologically meaningful?

3. Is the data generalizable?

- Were the groups that were studied sufficiently different from those in a general population so as to render the study data not applicable? For example, was the study population different from the general population in terms of age, socioeconomic status and environmental factors?

4. Does the association represent a cause-and-effect relationship?

- Is there sufficient evidence to satisfy convincingly the Bradford-Hill criteria assuming that the strength of the association is satisfied through the epidemiological data? In other words, does the evidence satisfy considerations of temporality, consistency, biologic plausibility, existence of alternate explanations for the association, dose response and specificity?

EFFECTIVE USE OF EPIDEMIOLOGY AT TRIAL

Understanding epidemiological principles will measurably aid defense counsel's ability to create an effective challenge to plaintiffs' causation testimony under *Daubert* and, if necessary, to convince jurors of the unreliability of a plaintiff's causation testimony not based on the proper use of epidemiology. Part A discusses increasing judicial awareness of epidemiology and a willingness to place

limitations on plaintiffs' causation evidence inconsistent with epidemiology. Part B offers some practice pointers for cross-examining plaintiffs' causation experts and conducting a direct examination of an epidemiologist.

A. Daubert Challenges to Causation Testimony Using Epidemiology

The Agent Orange and Bendectin litigation helped to increase judicial recognition and acceptance of the critical role that epidemiology plays in establishing causation. In the Agent Orange litigation, Judge Weinstein recognized that in a mass tort case, "epidemiological studies on causation assume a role of critical importance".[61] A number of courts have agreed with Judge Weinstein. For example, the Fifth Circuit in *Brock v. Merrell Dow Pharmaceuticals Inc.*[62] reversed the trial court's denial of a judgment notwithstanding the verdict and held that "the most useful and conclusive type of evidence in a case such as this is epidemiological studies".

When a plaintiff's expert relies exclusively on case reports, case series, animal or in vitro data, an aggressive *Daubert* challenge should be made to the admissibility of the causation evidence. Plaintiffs will argue that they should not be precluded from presenting their proof if no epidemiological evidence exists. In these cases, courts have been reluctant to exclude the plaintiff's evidence, although no logical reason exists under *Daubert* to permit a plaintiff to present proof of causation with speculative evidence.[63]

In most cases, there will be some epidemiological evidence. Circumstances include instances in which several studies uniformly fail to find an association, cases where plaintiffs' experts re-interpret negative studies in order to conclude an association exists, and those in which the study authors report having found an association. In each case, a *Daubert* challenge should be considered.

A number of courts have held that when the relative risk of a study is less than two, it is more likely than not that background causes produced the disease. In such cases, epidemiological evidence of the association is inadmissible and/or insufficient to support a finding of causation.[64]

A *Daubert* challenge also should be considered in those instances in which a plaintiff's expert has re-interpreted a negative epidemiological study to conclude that a statistically significant increased relative risk exists with respect to the exposure and disease. The Ninth Circuit recently recognized in *Lust v.*

---

61. In re Agent Orange Prod. Liab. Litig., 611 F.Supp. 1223, 1239 (E.D. N.Y. 1985), *aff'd*, 818 F.2d 187 (2d Cir. 1987), *cert. denied*, 487 U.S. 1234 (1988).

62. 874 F.2d 307, 311 (5th Cir. 1989), *modified*, 884 F.2d 166 (5th Cir. 1989) (per curiam), *cert. denied*, 494 U.S. 1046 (1990). *See also* Wade-Greaux v. Whitehall Laboratories Inc., 874 F.Supp. 1441, 1483 (D. V.I. 1994), *aff'd*, 46 F.3d 1120 (3d Cir. 1994) (since human experimental studies are unethical, epidemiology provides only evidence reasonably conclusive about associations that may occur in humans); Porter v. Whitehall Laboratories Inc., 791 F.Supp. 1335, 1347 (S.D. Ind. 1992), *aff'd*, 9 F.3d 607 (7th Cir. 1993); Thomas v. Hoffmann-LaRoche Inc., 731 F.Supp. 224, 226 (N.D. Miss. 1989) (while experts testified that ingestion of Accutane caused plaintiffs' seizures, they lacked epidemiological data or studies to support their opinions).

63. *See, e.g.*, Bowers v. Northern Telecom Inc., 905 F.Supp. 1004, 1010 (N.D. Fla. 1995) ("a cause-effect relationship need not be clearly established by . . . epidemiological studies before a doctor can testify that, in his opinion, such a relationship exists") (citations omitted).

64. Daubert, 43 F.3d at 1320 (relative risk of less than two tends to disprove legal causation); Marder v. G.D. Searle & Co., 630 F.Supp. 1087, 1092-95 (D. Md. 1986), *aff'd*, 814 F.2d 655 (4th Cir. 1987) ("neither of these risk figures indicate a 2-fold risk, and the 1.3 figure is well below the 2-fold risk level . . . The upper range of the confidence interval signifies the outer realm of possibilities, and the plaintiffs cannot reasonably rely on these numbers as evidence of the probability of the greater than 2-fold risk . . . In conclusion, plaintiffs fail to present sufficient evidence of causation and a verdict must be entered in favor of defendant.").

Merrell-Dow Pharmaceuticals Inc. that:

> When a scientist claims to rely on a method practiced by most scientists, yet presents conclusions that are shared by no other scientist, the district court should be wary that the method has not been faithfully applied. It is the proponent of the expert who has the burden of proving admissibility. To enforce this burden, the district court can exclude the opinion if the expert fails to identify and defend the reasons that his conclusions are anomalous.[65]

In those instances in which epidemiological studies have concluded that a relative risk of greater than two exists at a statistically significant level, a *Daubert* challenge may still be appropriate, because an epidemiological association does not, in and of itself, prove causation. Several courts have recognized that factors such as the Bradford-Hill criteria also must be met in order to prove causation.[66]

Finally, a *Daubert* challenge may be pursued when the proffered epidemiological studies lack adequate controls. Courts have excluded epidemiological studies when defendants can show that the epidemiological studies lacked adequate controls.[67]

B. Practice Pointers on Using Epidemiology at Trial

1. Cross-examining Plaintiffs' Experts

If epidemiological studies have not been conducted on the exposure that is the basis of the lawsuit and if a *Daubert* challenge fails, plaintiffs' experts will offer an opinion on causation based on their clinical experience and animal or in vitro studies. One way to cross-examine them is to review the fundamental difference between case reports, case series and epidemiology. Basic texts can be used to impeach witnesses if they refuse to admit that case reports, which form the basis of their opinions, are not an appropriate ground from which scientists reach conclusions on causation.

The cross-examination can begin by discussing general epidemiological principles and then lead to examples of when and why reliance on case reports can and often does produce erroneous conclusions regarding causation. Defense counsel should take the witnesses' primary strength (*e.g.*, clinical experience) and turn it into a liability by demonstrating that treating several patients with the same illness, who in many instances have been referred to the expert by counsel, constitutes referral bias.

The cross examination also should also attack the experts' testimony that they are called on "each day" when performing a "differential diagnosis" to determine causation. The impression plaintiffs' counsel wish to leave with jurors is that, in the "real world," treating doctors (like the expert) have to deal with problems of causation without the assistance of epidemiological studies.

One of the fallacies in this argument is a misunderstanding of the differential diag-

---

65. 89 F.3d 594, 598 (9th Cir. 1996).

66. *See* Claar v. Burlington N. R.R. Co., 29 F.3d 499, 501 (9th Cir. 1994) (expert opinions on specific causation that fail to rule out other causes are nothing more than mere speculation); Joseph Sanders, *The Bendectin Litigation: A Case Study in the Life Cycle of Mass Torts*, 43 HASTINGS L.J. 301, 329 (1992) ("Because we often may not understand the mechanisms by which a drug causes human injury, epidemiological correlation alone does not constitute proof of causation. Epidemiologists have developed a set of additional criteria to assist in determining whether the injury is the result of exposure to a particular product.").

67. *See, e.g.*, Muzzey v. Kerr-McGee Chemical Corp., 921 F.Supp. 511, 519-20 (N.D. Ill. 1996) (excluding causation evidence where experts relied on case reports and failed to exclude other potential causes); Grimes v. Hoffmann-LaRoche Inc., 907 F.Supp. 33, 35 n.2 (D. N.H. 1995) ("No epidemiological studies have been done which establish any relationship between Accutane and cataracts, and [plaintiff's causation expert] did not contend that causation can be proved by anecdotal evidence alone.").

There are undoubtedly some instances in which the treating physician needs to know the cause of a disease in order to treat the symptoms appropriately. In most cases, however, the process of conducting a differential diagnosis is limited to determining which, among several disease possibilities, the plaintiff actually has. This process does not generally require the physician to form conclusions regarding the etiologic agent for the suspected disease. For example, stomach cancer and ulcer may be among the differential diagnosis for stomach pain. In choosing between the two diagnoses, a physician need not determine what caused either the ulcer or cancer. Therefore, it is incorrect to allow plaintiffs' experts to testify unchallenged that a differential diagnosis requires a determination of causation.

Alternatively, situations occur where epidemiological studies are conducted that either because of a small sample size or because of the low incidence of the disease fail to demonstrate an increased relative risk at a level that is statistically significant. Plaintiffs' experts circumvent this problem by performing a "meta-analysis" of the studies.[68] When individual epidemiological studies are not large enough to reach statistical significance, epidemiologists under some circumstances can legitimately combine the data of several studies to determine whether the increased sample size is sufficiently large to generate statistically significant data.

Meta-analysis generally has been limited to randomly controlled experimental studies. Epidemiologists are beginning, however, to use this tool more frequently in analyzing the results of observational studies. Several courts have recognized that meta-analysis is a permissible analytical tool by which to evaluate epidemiological data.[69]

There are limitations to the proper use of meta-analysis. Use should be limited to those instances in which the data are homogenous.[70] If meta-analysis has been used to evaluate data that are not homogenous or data generated by studies whose designs are fundamentally different or which are subject to different biases, a *Daubert* challenge should be made to the testimony.

Plaintiffs' epidemiologists may concede that a statistical association does not alone prove causation. Many will agree that reaching an opinion on causation is the end result of an analysis of different types of scientific evidence. Depending on the strength, quality and availability of other such scientific data, defense counsel need to decide which of the Bradford-Hill criteria are most relevant to challenge the causation opinion drawn from the epidemiological evidence. Factors that are most understandable by juries include the inconsistency of results, alternative explanations, dose response and biologic implausibility.

In situations like the silicone breast litigation, where the epidemiological evidence refutes the plaintiffs' claims of causation, defense counsel must be prepared to address different arguments. Virtually no study can demonstrate a relative risk of one, unless there is a protective effect between the exposure and disease. As a result, plaintiffs' epidemiologists typically argue that the negative studies do not "prove" that the product is safe or that it could not have been the cause of the injuries.

This argument must be attacked on cross-examination and in the defense case by explaining that in science, as in law, it is the burden of the party asserting a proposition to prove the claim. Because of

---

68. A meta-analysis is defined as "the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings," Gordis, *supra* note 4, at 253 (quoting G. Glass, *Primary, Secondary and Meta-Analysis of Research*, 5 EDUC. RES. 3-8 (1976)).

69. *See e.g.*, Tobin v. Astra Pharmaceutical Prod. Inc., 993 F.2d 528, 538-39 (6th Cir. 1992); In re Paoli R.R. Yard PCB Litig., 916 F.2d 829, 856-57 (3d Cir. 1990), *cert. denied*, 499 U.S. 461 (1991).

70. Gordis, *supra* note 4, at 253.

the role of chance, it is very difficult to prove a negative proposition scientifically. For example, bringing back rocks from the moon would be highly probative evidence to refute a claim that the moon is made of cheese. However, plaintiffs' experts would argue that the moon rocks had not disproved the claim because the correct area of the moon had not been searched that "certainly" would have proved that the moon is made of cheese.

Although jurors will be instructed that it is the plaintiff's burden to prove claims of defectiveness, some jurors will nevertheless feel that it is the manufacturer's responsibility to prove that its product is safe. For these jurors, it will be difficult for them not to expect or require defense experts to disprove what the plaintiffs' experts have asserted. This can be particularly difficult when dealing with epidemiological evidence. As explained earlier, the null hypothesis of an epidemiological study, as generally stated, is that there is no association between exposure and disease. Plaintiffs' experts will argue that in light of the hypothesis, the "clear" purpose of such a study was to prove that no association exists (*e.g.*, the product is safe). If the study fails to show a RR=1 at levels of statistical significance, the expert will argue that the study failed to prove the product was safe.

Another approach used by plaintiffs' epidemiologists is to calculate the power of the study *after* it has been completed and to argue that the study was not strong enough to find the association that they are certain exists. As discussed earlier, *post hoc* power calculations are an inappropriate basis to criticize data after the study has been completed. Such testimony should be challenged under *Daubert*.

Another argument used by plaintiffs confuses the Alpha level and the legal burden of persuasion. For example, if a study shows an increased RR but not at a level of statistical significance, plaintiffs' experts typically argue that although the study may not "reach the scientific level of 95 percent certainty," the evidence is nevertheless reliable and can form the basis of the experts' conclusion that at a level of 51 percent (*e.g.*, preponderance of the evidence), the evidence proves that exposure caused the disease.

Such an argument misrepresents the role of statistical significance in evaluating scientific data and reflects a misunderstanding of the legal burden of persuasion. The major thrust of *Daubert* is a recognition that juries should not be exposed to unsubstantiated and unreliable scientific speculation simply because a witness says that an opinion has been formed to a "reasonable degree of scientific probability." In the scientific world, before data are recognized as having been scientifically "proved," it must be demonstrated (assuming the experiment was done correctly) that there was no more than a 1 in 20 probability that the results obtained could have been obtained by chance alone. A scientist departs from accepted scientific methodology by telling a jury that a scientific "fact" has been proved simply because the legal standard is 51 percent. Such testimony is based on speculation and ought not be admissible.

Proper scientific methodology requires that conclusions be drawn from the aggregation and analysis of statistically significant data. Only such data can be said to be "proved." Experts may disagree on the proper interpretation of agreed-on scientific facts, and those opinions can be presented in the courtroom. But before an expert can incorporate or base an opinion on a scientific "fact," that fact must be proved, as required by scientific methodology, to a degree of statistical significance. The opinion drawn from an analysis of the scientifically proved facts is one courts require to be expressed to a degree of reasonable probability.

While it is entirely appropriate for an expert to express an opinion to a reasonable degree of probability in a courtroom, such an opinion must be predicated on scientific evidence that has been developed in accor-

dance with appropriate scientific methodology. It is the sum total of the expert's opinion, plus all other evidence that is evaluated by the trier of fact by the preponderance of the evidence standard. Therefore, plaintiffs' experts should not be permitted to assert that data that have not been demonstrated to be statistically significant nevertheless have been proved causation at the lower "legal" standard of 51 percent.[71]

Another technique used by plaintiffs' experts to minimize the significance of negative epidemiological data is to suggest that a different result would have been achieved if small numbers of exposure or disease cases had been added to the study. In these cases, the experts argue that not all disease cases or exposure cases were included in the study and may even point to the plaintiff as an example that the study was "incomplete." They will argue that by adding only a few additional cases of disease or exposure, depending on whether the study is a case-control or cohort study, the result would be a statistically significant increased RR.

Despite the simplistic and appealing nature of this argument, it is fallacious because the hypothetical treats the controls differently. If similar assumptions are made about the control group, then the RR will not change. Defense counsel can refute the argument by explaining that because an epidemiological study is a sample of a larger population, not all cases are included. More important, there is generally an equal reason to assume that additional non-disease controls were excluded who had also been exposed. Unless this assumption is shown by the plaintiff to be incorrect, adding non-diseased controls who had been exposed will negate the effect of adding additional exposed disease cases.

2. Epidemiological Evidence on Direct Examination

If a *Daubert* challenge to a plaintiff's causation expert has failed, the defendant is faced with presenting epidemiological evidence after the jury has heard causation testimony from physicians who describe themselves as "treating physicians." The impression that plaintiffs' experts seek to leave with the jury is that they are in the best position to render an opinion on causation because they have purportedly "treated" (generally a perfunctory examination) the plaintiff and rendered a "differential diagnosis." Often, in anticipation of the defense presenting epidemiological evidence, the jury has also been told to distrust the epidemiological evidence because any evidence is capable of statistical manipulation.

It is in this context that the defense counsel must convince the jury that the "treating" physicians' causation opinion does not have scientific validity and is wrong. How is this accomplished?

First, after an epidemiologist explains what epidemiology is, but before the witness offers a detailed discussion of how an epidemiological study is performed, the jury should be told about the critical importance of epidemiology. The jury should be told that most of the advances in medical science in the 20th century with respect to improving public health, identifying causes of disease, developing treatments and preventative measures are the result of epidemiological studies. Once the jury has an appreciation of epidemiology in the public health context, the testimony may focus on why controls are critical before judgments on causation are made. Examples should be offered of how physicians have mistreated diseases in the past based on medical judgments formed from uncontrolled case reports. This helps the jury understand the critical role of epidemiology in clinical practice.

For example, it was through the use of control populations and record keeping, tools essential to epidemiological investigations, that physicians learned that treat-

71. *See generally* Reference Manual, *supra* note 8, at 153.

ments of a variety of ailments—such as bloodletting—were no more efficacious than no treatment at all. For several hundred years, physicians believed that bloodletting was an effective treatment for the common cold because the "treating" physicians observed that the cold would get better within five to seven days, unless, of course, the patient died of the complications of bloodletting. It was not until physicians understood the critical importance of a control population that they were able to learn that untreated patients with similar colds also got better in five to seven days.

During this discussion, it bears repeating that although case reports are important for generating hypotheses of possible causal relationships, anecdotal evidence derived from case reports is not a scientific basis from which one can reach conclusions regarding causation.

Next, the defense must explain how epidemiological studies are conducted. If both cohort and case-control studies have been conducted on the product, the expert needs to explain the difference between cohort and case-controlled studies and the unique value of each.

The first step in helping juries feel comfortable with statistical data in an epidemiology study begins with helping the jurors understand that the disease or symptoms experienced by the plaintiff also occur in the general population that has not been exposed. Once the jury understands that the disease or symptoms occur in the non-exposed population, it is easier for the epidemiologist to explain and the jury to accept the statistical conclusions drawn from the data.

The epidemiologist needs to use common sense examples to explain the difference between a significant and insignificant statistical difference. Baseball averages are one example. Coin tossing is another. The expert also needs to explain the significance of a negative epidemiological study. Because of the essential importance of this data, they must be presented as aggressively as one can consistent with the data. For example, one often hears epidemiologists state that a negative epidemiological study "does not support the plaintiff's arguments." While certainly true, this characterization does not maximize the value of the data. For example, epidemiological data generated on Bendectin and silicone breast implants have consistently demonstrated the lack of an association between exposure and disease. The uniformity of the data is itself a powerful piece of affirmative evidence that refutes the plaintiffs' claims.

If the defense epidemiologist is called on to criticize an epidemiological study, the epidemiologist will need to explain bias and confounding factors. Again, the key to explaining these concepts lies with the ability of the witness to use examples which are understandable to the jury. The association between infant shoe size and improved vocabulary is a good example. Most jurors would probably understand that the confounding variable that drives this association is not the increased shoe size, but age. This simple example helps the epidemiologist explain the effect which confounding variables have in a study which does not control for the variables.

CONCLUSION

Defending drug and medical device cases from claims that the product caused a systemic chronic disease requires counsel to understand the strengths and limitations of epidemiology. Courts are becoming increasingly willing to strike testimony that is inconsistent with epidemiological data. The challenge to defense counsel is to present epidemiology in a way that helps the jury understand that disease causation is determined through epidemiology and not reliance on case reports by physicians performing "differential diagnosis."

ADDENDUM

After this article was ready for publication, the U.S. Supreme Court decided *General Electric Co. v. Joiner,* 118 S.Ct. 512

(December 15, 1997). *Joiner* re-emphasizes that a trial court is required, as part of its gatekeeping role, to evaluate not only the methodology used by an expert, but also whether the expert's conclusion meet *Daubert* standards. The "abuse of discretion" standard of review, as adopted in *Joiner,* makes it imperative that defense counsel thoroughly educate the trial judge on the important role of epidemiology in establishing medical causation.