REPRINT

COUNTERPOINT®

AN OFFICIAL PUBLICATION OF THE PENNSYLVANIA DEFENSE INSTITUTE
An Association of Defense Lawyers and Claims Executives

JANUARY 2000

Evaluating Expert Testimony With Respect to Medical Causation: The Use of Epidemiological Studies in Court

By: Madeline M. Sherry, Esquire and Frederick E. Blakelock, Esquire
Hecker Brown Sherry and Johnson, Philadelphia, PA

In toxic tort cases, medical causation is often the plaintiff's most difficult element of proof. The first hurdle a plaintiff faces is proving general causation, that the substance at issue is capable of causing the disease or condition from which the plaintiff suffers. Special causation is generally more challenging to prove. The plaintiff must show that the suspected toxin at issue more likely than not was a substantial contributing factor in causing his or her particular injury.

Expert testimony is the tool with which plaintiffs meet their burden of proving causation. One of the most common sources of information expert witnesses may rely on to support their opinions regarding causation are epidemiological studies. While epidemiological studies can be useful to assist a scientist in determining whether it is likely that a suspected toxin might have the ability to cause a particular condition, a problem arises when experts rely too heavily on epidemiological studies to prove specific causation. In many cases, such opinions should be viewed with skepticism.

Blum v. Merrell Dow Pharmaceuticals, Inc.

On October 18, 1999, the Pennsylvania Supreme Court heard oral argument in the appeal of the Superior Court's decision in *Blum v. Merrell Dow Pharmaceuticals, Inc.*, 705 A.2d 1314 (Pa. Super. 1997). At issue is the opinion of the plaintiff's expert that Bendectin, the anti-nausea drug ingested by plaintiff's mother during pregnancy, caused plaintiff to be born with clubfeet. At trial, plaintiff offered the testimony of four experts, all of whom opined that Bendectin was a human teratogen (i.e., had the ability to cause development malformations). However, only one expert, Dr. Alan K. Done, testified on the issue of specific causation: that Bendectin caused the particular plaintiff's clubfeet. In support of his opinion, Dr. Done relied on four common methods of scientific analysis to determine whether a given substance may have harmful propensities: chemical structure analysis examining the molecular structure of doxylamine (the allegedly harmful ingredient in Bendectin); in vivo studies of live animals; in vitro studies of animal cells; and epidemiological studies. Dr. Done conceded that the chemical structure analysis, in vivo studies and in vitro studies did not prove to reasonable degree of certainty that Bendectin was a human teratogen, much less that Bendectin had specially caused the plaintiff's club feet. The crux of Dr. Done's opinion on causation, therefore, was his reliance on his own recalculation of data in a previously published epidemiological study that found no statistically significant correlation between Bendectin and club feet (or other birth defects).[1]

Pennsylvania currently adheres to the "general acceptance" standard set out in *Frye v. United States*, 293 F. 1013 (D.C. Circ. 1923). In Pennsylvania, "in order for scientific testimony indicating that an event causes a particular result to be admitted, there must be a showing, not that the studies establishing the causal relationship follow generally accepted methodologies, but that the existence of the causal relationship is generally accepted by the relevant medical community." *McKenzie v. Westinghouse Elec. Corp.*, 674 A.2d 1167, 1172 (Pa. Cmwlth. 1996).

In 1993, the *Frye* standard was superceded in federal courts by the standard set out in *Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 579, 113 S. Ct. 2786, 125 L. Ed. 2d 469 (1993). In *Daubert*, the United States Supreme Court held that general acceptance of the principle underlying an expert's opinion is not necessary; rather, trial courts must determine whether the expert's opinions are based on sound scientific methodology. In 1998, Pennsylvania adopted rules of evidence, including rules 701 through 706 involving opinion and expert testimony. Although the Pennsylvania Rules of Evidence involving expert testimony are similar to the Federal Rules of Evidence dealing with expert testimony, the Pennsylvania Supreme Court has not yet decided whether *Daubert* applies to experts in Pennsylvania. While the federal standard requires only that the expert's conclusion be based on reliable methodology, in *Joiner v. General Electric Company*, 522 U.S. 136, 118 S.Ct. 512, 139 L.Ed. 2d 508 (1997), the United States Supreme Court opined:

> . . .conclusions and methodology are not entirely distinct from one another. Trained experts commonly extrapolate from existing data. But nothing in either *Daubert* or the Federal Rules of Evidence requires a district court to admit opinion evidence which is connected to existing data only by the ipse dixit of the expert. A court may conclude that there is simply too great an analytical gap between the data and the opinion proffered."

Id., 522 U.S. at 146.

The Superior Court in *Blum v. Merrell Dow Pharmaceuticals*, 705 A.2d 1314 (Pa. Super. 1997), came to a similar conclusion. The court determined that Dr.

Done's opinion was supported neither by a generally accepted scientific principle *(Frye)* nor by reliable methodology *(Daubert)*. *Id* at 1321. The court found that Dr. Done's methodology was unreliable because he failed to "standardize" the raw data from the previously published study that he used to draw his conclusion regarding causation. Specifically, Dr. Done failed to account for "selection bias" (occurring when the population contained in the study is not representative of the general population) and "confounding bias" (occurring when the study fails to control adequately for certain factors that vary between groups so that the populations being compared are not truly similar). *Id.* at 1324. The court concluded:

> Although the general types of studies relied on by the Blums' experts are universally accepted as good science, the way they have utilized them to draw conclusions is not . . . Dr. Done's elimination of standardization from the epidemiological analysis made the epidemiological methodology not generally accepted. To be more specific, his epidemiological analysis was so flawed as to render his conclusions unreliable and therefore inadmissible.

Thus, the resolution of *Blum* may not be dependent on what the Pennsylvania Supreme Court decides is the appropriate standard for expert testimony in Pennsylvania, since the court may determine that Dr. Done's reliance on a flawed epidemiological analysis renders his opinion inadmissible regardless of the standard by which it is reviewed.

As *Blum* demonstrates, epidemiology, like many statistics-based disciplines, is open to abuse.[2] Therefore, it is important for courts and counsel to understand the scientific method by which epidemiological data should be evaluated, so that unreliable opinions based on unreliable methodologies are not permitted to confuse the jury.

The Science of Epidemiology

Epidemiology is the study of patterns of disease across human populations. Because epidemiological studies are complex and require great care in both compilation and interpretation of data, the usefulness and any epidemiological study depends on the quality of the underlying data, the reliability of the methodology and the validity of the interpretation. Dore, Michael, "A Commentary on the Use of Epidemiological Evidence in demonstrating Cause-in-Fact," 7 *Harvard Environmental Law Review,* 429, 432 (1983). While epidemiology can be a useful tool to assist a properly trained scientist in drawing conclusions about causation, epidemiological studies should not form the sole basis for conclusions about medical causation in individual cases, since they cannot prove causation in any particular case. *Id.* at 433-434. Commentators have suggested that by allowing epidemiological evidence to be used as evidence of individual causation, some courts have demonstrated a lack of understanding of the principles underlying the science of epidemiology. *Id.* at 435. Studies show only the relative risk that a given substance posed to the members of a given population. While the risk applies to the individual plaintiff, evidence of risk is not evidence of causation. *Id.* at 436-437. Thus, an epidemiological study must be used as but one part of a valid scientific methodology, regardless of the strength of the statistical association.

Evaluating Causation Opinions Based on Epidemiology.

In 1965, Sir Austin Bradford Hill, Professor Emeritus of Medical Statistics at the University of London, outlined criteria that should be used in assessing the likelihood of a causal relationship of a given substance and a disease, based on epidemiological data showing an association between the two factors. Hill, AB: The environment and disease: association or causation? *Proc R Soc Med* 58:295-300; 1965. Those criteria are: (1) the strength of the association; (2) the consistency of the association; (3) specificity of the association; (4) temporally correct relationship; (5) biological gradient (dose-response relationship); (6) biological plausibility; (7) coherence; (8) experiment; and (9) analogy. Although Hill was careful of point out that these are guidelines, not strict requirements, *id* at 299, these criteria are still used to assess the validity of conclusions drawn from epidemiology, and may be useful to counsel when evaluating expert opinions or taking expert depositions.

"Strength of association"

"Association" is defined as "the statistical dependence between two variables, that is, the degree to which the rate of disease in persons with a specific exposure is either higher or lower than the rate of disease among those without that exposure." Cetrulo at 5-41, citing C.H. Hennekens & J.B. Buring, *Epidemiology in Medicine.* 122 (S.L. Mayre at ed. 1987). Generally, the stronger the statistical association, the more likely that it represents a cause and effect relationship. The finding of an association between a substance and a disease, however, is simply a statistical relationship between the substance and the disease, and is only one piece of information that may be used to establish a cause and effect relationship. Cetrulo, Lawrence, *Toxic Torts,* Vol. 1, Chapter 5, 1993 at 5-41.

"Relative risk"

The strength of an association is usually measured by "relative risk" or the ratio of the disease rate in those exposed to the suspect toxin to the rate in those who have not been so exposed. Garry D. Friedman, *Primer of Epidemiology* 1 (3rd ed. 1987) at 183. If there is no association between the suspect factor and the disease, the relative risk is expressed as 1.0, indicating that the incidence rates for the exposed individuals and the non-exposed groups are equal. Bert Black and Albert Lilienfeld, *Epidemiologic Proof in Toxic Tort Litigation,* 52 Fordham L. Rev. 732, 758 (1984). For example, the relative risk of lung cancer in cigarette smokers as compared to nonsmokers is approximately 10:1, whereas the relative risk of pancreatic cancer is about 2:1, suggesting that cigarette smoking is more likely to be a causal factor in the development of lung cancer than for pancreatic cancer. Friedman at 183. It is rare to observe a relative risk above 10. At such a level, due to the strength of the association, one can be reasonably certain that the association represents a causal relationship. Black and Lilienfeld at 758.

Commentators and some courts have suggested that a relative risk of 2.0 should be used as a yardstick (in conjunction with other evidence) to establish admissibility on causation, and that studies establishing a relative risk of less than 2.0 (i.e. double the risk of the unexposed group) should be not be admissible on the issue of causation. Cetrulo at 5-44.

"Attributable risk"

Another statistical measure of the strength of association is the "attributable risk", or the proportion of the disease in a given population that is statistically attributable to the exposure at issue. Cetrulo at 5-46. For example, in 1984 the risk of lung cancer attributable to smoking was approximately 80%. Black and Lilienfeld at 760. Thus, based on this attributable risk figure, it might be argued that if smoking were eliminated in the United States, the incidence of lung cancer would decline by approximately 80%. *Id.*

"Standardized mortality ratio"

One other common statistical measure of the strength of an association between exposure and disease is the Standardized Mortality Ratio ("SMR"). A standardized mortality ratio is defined as the ratio of the number of deaths observed in a study to the number of deaths expected in the general population, multiplied by 100. Cetrulo at 5-46. An SMR greater than 100 percent for a particular exposed population reflects a positive association between exposure and disease. The higher the SMR for the particular exposed population, the greater the strength of the association between exposure and disease. Cetrulo at 5-47.

"Confidence Intervals"

When assessing the strength of an association, it is important to be aware of the reported confidence interval of the study. A confidence interval identifies a range within which the true risk probably lies. Cetrulo at 5-51. Confidence intervals are generally calculated to a 95% degree of confidence. *Id.* If the confidence interval includes relative risk values of 1.0 of less, the study is not considered to have shown a statistically significant association, since a relative risk of 1.0 indicates the exposed population is not any more likely to develop the disease than the unexposed population. Hence, epidemiologist do not consider a relative risk with a 95% confidence interval that includes values of 1.0 or less, to be statistically significant evidence of a positive association between the studied factors. *Id.* at 5-52.

"Consistency of association"

To be reliable, epidemiological studies should show a consistent level of associated risk over and over again across different populations exposed to the same factor. That is, the question is whether the association has "been repeatedly observed by different persons, in different places, circumstances and times?" Hill at 296. For example, if one study showed a very strong association between smoking and lung cancer, and another study, utilizing a different population, demonstrated a weak association, an expert would pause before accepting wholeheartedly the association in either study and before relying on just one of the studies in support of his or her opinion. Moreover, the results of the study should be reproducible by other scientists studying the same factor. Results that cannot be replicated are not considered reliable.

"Specificity of association"

It is also important to determine the extent to which the association is limited specific types of individuals and to particular circumstances and types of diseases. Hill at 297. The greater the specificity of the association, the more likely that the association represents a causal relationship.

"Temporally correct relationship"

Scientists must establish whether a certain factor causes a disease or whether a disease causes a certain factor. In other words, "does a particular diet lead to disease or do the early stages of the disease lead to those particular dietetic habits?" Hill at 297.

In addition, in order to support evidence of causation, the disease should manifest itself in a fairly consistent manner after exposure to the suspect factor. For example, if studying the relationship of smoking to lung cancer, one should expect a fairly consistent latency period for lung cancer at a given level of smoking. If some cancers appeared in seven years and others in forty years, the conclusion that smoking was the cause should not be readily accepted. In *In Re: TMI Litigation,* 1999, U.S. App. Lexis 28415, *300-*302 (3rd Cir. 1999), the court explained:

> The reactor accident occurred in 1979 and the last year covered by [the study at issue] was 1985. Therefore, the period between the accident and the last year of the study was six years. Consequently, the ability to make a plausible association between the accident and a diagnosis of post-accident lung cancer depends upon the length of the latency period. If the latency period is four to eight years, as [the expert] claimed, then a sufficient latency period elapsed between exposure and diagnosis to make a plausible association between the exposure and the lung cancer. If, however, the latency period is ten to fifteen years, as defendants claim, then an insufficient latency period elapsed between the date of the accident and the last year of the study to draw a plausible association between radiation released by the reactor accident and a diagnosis of lung cancer. *Id.*

Biological gradient (dose-response relationship)"

A clear does-response curve lends strength to any causation theory. Hill at 298. Generally, increased dosage of the suspected toxin should increase the incidence or severity of the disease.

Moreover, the level of exposure to the suspect factor must be at a level sufficient to cause the condition. All chemicals, even everyday benign substances, can produce an adverse effect at some level and duration of exposure. Harbison, Raymond, *Hamilton and Hardy's Industrial Toxicology,* 5th ed.; 98 at p. 4. Accordingly, risk is determined by evaluating the exposure required to produce toxicity. The basis for risk assessment is identifying those circumstances and conditions under which an adverse effect can be produced. *Id.* At a minimum, in order to provide a reasonable basis for an opinion on causation, the expert should have specific information as to the level of exposure of the studies population.

"Biological plausibility"

Simply stated, any opinion on causation should make sense. For example, since asbestos fibers are inhaled into the lungs, it is plausible that asbestos could cause harm to the lungs. To the contrary, an expert who opines that asbestos causes prostate cancer espouses a less plausible theory, since it would be difficult for asbestos fibers to travel to the prostate.

Moreover, in addition to being plausible, the cause and effect theory should be generally consistent with established scientific principles and knowledge (i.e., scientific "coherence"). Hill at 298.

Other Methodological Considerations

Other methodological considerations that should be accounted for include whether:

- the study encompassed a population large enough to reflect the characteristics of the total population of interest;
- the study identified a control population as a clear norm from which to measure deviations;
- the deviations from the norm are large enough to be distinguished from random fluctions; and
- the study accounted for a wide range of potentially pertinent factors (e.g. age, race, sex, occupation, geographic location and everyday exposure to hazardous materials).

Dore at 432.

These factors are important so that the study is "generalizable" from the studied population to society at large, or more specifically, the plaintiff at issue. For example, Dr. Irving J. Selikoff studied the health effects of asbestos on insulators. *See, e.g.,* Selikoff, Irving J., et. al., "Mortality Experience of Insulation Workers in the United States and Canada, 1943-1976", *Annals New York Academy of Sciences,* (1979). Dr. Selikoff's articles drawing conclusions based on these studies are widely accepted as authoritative. Because Dr. Selikoff studied insulators who had significant exposure to asbestos dust on a daily basis, however, the extent to which these conclusions regarding the relative risk of certain diseases from asbestos exposure can be applied to individuals not as heavily exposed to asbestos, such as painters or carpenters, is not clear.

In *In Re: TMI Litigation,* 1999 U.S. App. Lexis 28415, *285-286 (3rd Cir. 1999), the court affirmed the district court's preclusion of an expert's opinion that was based on an epidemiological study in which the method used for choosing the participants was not controlled so as to minimize selection bias:

> The absence of evidence that the Aamodts selected the participants in the groups in a manner consistent with that suggested by the Reference Manual on Scientific Evidence creates a profound flaw in the [the expert's] methodology . . . An epidemiological opinion based on such a study is not reliable, and the District Court did not abuse its discretion when it excluded [the expert's] testimony. *Id.*

Moreover, in order to be part of a valid scientific methodology, the expert must examine other criteria apart from the epidemiological study, such as experimental evidence in animals, the confounding risk factors, and the coherence and weight of the evidence. Harbison at 5. Thus, any investigation into an expert's opinion as to causality should center on the reliability of the study and the extent to which the study is a part of valid scientific reasoning.

Experimental Evidence in Animals

Animal studies are useful tools, but one cannot form a valid opinion about causation of disease in humans from animal studies alone. Harbison at 6. For example, animal studies can be used to determine whether a specific agent is a potential human carcinogen, but they cannot sya whether the chemical is an actual carcinogen. *Id.* Because of the important differences in the anatomy, physiology and biochemistry of humans and animals, extrapolation of results to humans should always be viewed with extreme caution in any causation analysis. *Id. See also: Blum v. Merrel Dow,* 705 A.2d at 1323 ("[A]nimal studies without epidemiological studies cannot prove causation in humans because drugs do not have the same effect on humans as they do on animals; the goes given to animals in animal studies are very different from those given to humans.")

Confounding Risk Factors

The question of possible alternative causes for any condition is especially vital in drawing conclusions with respect to the question of specific causation. For example, if an expert bases his opinion on causation on an epidemiological study that concludes the relative risk of colon cancer for a person heavily exposed to asbestos during his work is 1.6, the question of whether asbestos caused a particular plaintiff's colon cancer is dependent on an analysis of various other possible causes, in addition to a detailed analysis of the extent of the plaintiff's occupational asbestos exposure. A high fat diet, sedentary lifestyle, excessive alcohol use and genetics are all alternative causes of colon cancer. It would be inappropriate to draw any conclusions with respect to causation based solely on an epidemiological study, regardless of the strength of the association, without ruling out possible other causes of the condition. This is simply sound scientific reasoning.

For example, in *General Electric v Joiner,* 522 U.S. 136, 118 S. Ct. 512, 139 L.Ed. 2d 508 (1997), plaintiff alleged that he had developed small cell lung cancer from his occupational exposure to PCBs. The Court recognized the existence of potential other causes, noting that "[the plaintiff]" had been a smoker for approximately 8 years, his parents had both been smokers, and there was a history of lung cancer in his family. He was thus perhaps already at a heightened risk of development lung cancer eventually." *Id.,* 522 U.S. at 139. Thus, expert opinions should take into account other risk factors for the plaintiff's alleged disease.

Meta-Analysis

Rather than relying on individual epidemiological studies, some experts will perform what is known as "meta-analysis". Meta-analysis involves pooling data from a number of epidemiological studies (in order to enhance the sample size) and comparing the results of these pooled data with results produced by each study individually. *Hines v. Consolidated Rail Corp.,* 926 F.2d 262 (3rd Cir. 1991).

This technique can suffer from the same weaknesses as individual studies, and can be similarly abused. For example, in *In re: Paoli R.R. Yard PCB Litigation,* 916 F.2d 829 (3rd Cir. 1990), the plaintiff's expert opined that there was a causal connection between PCB exposure and human illness by utilizing meta-analysis. The expert claimed to have pooled the data from several studies, none of which had found such a causal connection. The district court found that this proffered testimony was unreliable. On appeal, the Third Circuit determined the district court's basis for exclusion of the expert's testimony was erroneous, since the court did not develop a sufficient record and make specific findings on reliability issues. *Id.* at 858. However, the Court of Appeals noted that the meta-analyses "are, at times, used in circumstances in which they should not be." *Id.* at 857.

Conclusion

When evaluating the reliability of epidemiological studies as evidence of cau-

sation, it is important that defense counsel have a basic understanding of the principles of epidemiology and it inherent limitations. Moreover, an awareness of epidemiology's place in an overall reliable scientific inquiry will become more fundamental should the Pennsylvania Supreme Court decide to follow the federal standard for admissibility of expert testimony.

ENDNOTES:

1. This testimony was similar to the expert testimony at issue in *Daubert v. Merrell Dow,* 509 U.S. 579, 113 S.Ct. 2786, 125 L.Ed. 2d 469 (1993). In *Daubert,* the Plaintiffs conceded that more than 30 published studies involoving over 130,000 patients had not found Benedectin to be a human teratogen. Howeever, plaintiff's experts based their opinion on in vitro and in vivo animal studies, pharmacological studies of the chemical structure of Benedectin, and "readings" of previuosly published epidemiological studies. *Id.*, 509 U.S. at 582-583.

2. *See also, In re: TMI Litigation,* 1999 U.S. App. LEXIS 28415, *282 (3rd Cir. November 2, 1999) ("[The plaintiff's expert] admits that 'statistical analysis of epidemiological data is much abused.")