

2003

The Legal System's use of Epidemiology

Arthur H. Bryant

Alexander A. Reinert

Benjamin N. Cardozo School of Law, areinert@yu.edu

Follow this and additional works at: <https://larc.cardozo.yu.edu/faculty-articles>



Part of the [Law Commons](#)

Recommended Citation

Arthur H. Bryant & Alexander A. Reinert, *The Legal System's use of Epidemiology*, 87 *Judicature* 12 (2003).

Available at: <https://larc.cardozo.yu.edu/faculty-articles/382>

This Article is brought to you for free and open access by the Faculty at LARC @ Cardozo Law. It has been accepted for inclusion in Articles by an authorized administrator of LARC @ Cardozo Law. For more information, please contact christine.george@yu.edu, carissa.vogel@yu.edu.

The LEGAL SYSTEM'S use of EPIDEMIOLOGY

by Arthur H. Bryant and Alexander A. Reinert

Both law and science are truth-seeking endeavors. In at least one respect, lawyers and scientists are like Agent Mulder on the X-Files: we believe that the truth is out there and our goal is to find it. This article is devoted to exploring and improving the means by which law relies on scientific disciplines, particularly epidemiology, to ascertain the truth.

While there are obvious differences between the processes used to search for truth in the scientific and legal arenas, the importance of science in the law is difficult to overstate. The law's interest in and reliance on science had been growing since well before the U.S. Supreme Court, in *Brown v. Board of Education* (1954), made use of social science research to reject the notion that racially separate education was really equal. Now, in courtrooms throughout the country, judges and juries look every day at scientific evidence to determine the truth (and to decide who wins and loses) in a wide array of contexts, including toxic tort, employment discrimination, environmental protec-


tion, products liability, civil rights, and criminal cases.

Unfortunately, although science and the law have similar truth-seeking goals, our judicial system has suffered from the failure to apply accurately scientific knowledge in the courtroom. Epidemiology (the branch of medicine dealing with the incidence and prevalence of disease in large populations) is a critical scientific discipline that has been misused by the law in the past 25 years. Some of this misuse is not the outcome of particular legal rules. Judges, lawyers, or witnesses who do not adequately understand epidemiology simply misapply it. Regretfully, however, judges and policy makers also have established systemic rules that give epidemiological studies either more or less weight than scientists would give them. As a result, the law is being taken further from the truth.

This article reviews six ways in which some courts are misusing epidemiological studies. These errors often preclude courts (and especially juries) from considering evidence that scientists (and especially epi-

demiologists) would readily consider. And, perhaps most galling, the justification given by judges for precluding reliance upon this expert testimony is that the testimony is not really "scientific." On this basis, the views of highly-accomplished scientists have been barred from the courts.¹ This article explains and offers some reasons for the errors, contends that epidemiological studies should be given the same weight and consideration in the legal arena that they are given in the scientific

This article is an abridged version of a presentation delivered at "From Epidemiology to Policy: A Symposium on the Translation of Epidemiologic Evidence into Public Health Policy" on July 19, 1998. The symposium was jointly sponsored by the Center for Epidemiology and Policy and the Risk Sciences and Public Policy Institute of Johns Hopkins School of Public Health and the American College of Preventive Medicine. A version of this paper was previously published in the *American Journal of Epidemiology*. Some citations have been updated to reflect legal developments since the presentation was delivered and since the *American Journal of Epidemiology* paper was published, although not all citations are included here. For a copy of the original paper, a copy of this article that includes all footnotes, or the *American Journal of Epidemiology* article, please contact Arthur Bryant at Trial Lawyers for Public Justice, Ordway Building, One Kaiser Plaza, Suite 275, Oakland, CA 94612-3684.



Both law and science are truth-seeking endeavors. Unfortunately, the judicial system has failed to apply accurately scientific knowledge, and particularly epidemiology, in the courtroom.



arena, and, finally, suggests several ways that epidemiologists, lawyers, judges, and other policy makers can work together toward this goal.

Isolated misuse

Jurors and judges in state and federal courts increasingly are called upon to evaluate the strength of a litigant's case based in part upon the presentation of epidemiological studies. Judges, whether making the initial determination of admissibility or adjudicating post-trial sufficiency of the evidence motions, must critically consider the nature of the evidence presented, including the role that epidemiology plays in informing scientific testimony. Thus, an initial problem for our justice system is the lack of scientific training of most federal and state judges. This lack of training makes judges vulnerable to misunderstanding and manipulation.

We briefly discuss here some examples of the difficulty judges have faced in correctly understanding concepts in epidemiology. Some of the examples reveal a basic unfamiliarity with general scientific principles, but are probably harmless. Take, for example, courts that have referred to "Koch's Postulates" as a means of evaluating epidemiological studies.² What these courts have in mind is not Koch's Postulates, but a modified version of Hill's suggestions for analyzing epidemiological studies.³ Fortunately, these courts have been right on the fundamentals, because Hill's criteria for evaluating epidemiological studies are generally accepted in the scientific community, and the error is understandable in part, given that the first edition of the Federal Judicial Center's *Reference Manual on Scientific Evidence*, upon which many courts rely, also misidentified Hill's criteria as Koch's Postulates.⁴

In some cases, however, the mistakes made by courts in evaluating epidemiologic evidence carry more substantive consequences. For example, the U.S. Court of Appeals for the Fifth Circuit, in *Brock v. Merrell Dow Pharmaceuticals* (1989), considering the sufficiency of scientific evidence

presented in a case involving an allegation that Bendectin caused birth defects, illustrated the difficulty some courts have understanding the proper interpretive weight to give confidence intervals.

First, the court asserted that the use of confidence intervals eliminated any need to analyze a particular study for recall bias or confounding. This is plainly false, because a confidence interval—which simply represents the range of values that, with a specified degree of certainty, is likely to contain the true measure of association between the disease and the proposed cause—has no connection to recall bias or confounding. In addition, the court stated that if a confidence interval included 1.0 in its range of possible values, then "no statistically significant conclusions could be drawn" from a study.⁵ This treatment of confidence intervals as interchangeable with "significance testing"—a proposition for which *Brock* has been cited by numerous courts—is a common misunderstanding, and will be discussed in detail in the section on structural misuse of epidemiology.

Similar examples abound. In a district court case, a federal judge starkly demonstrated the difficulty that some courts have with the basic statistical method of hypothesis testing. After noting that accepted p-values for hypothesis testing included "5%" and "1%," the court equated those percentiles with the numerical values 0.5 and 0.1, respectively, instead of the correct values of 0.05 and 0.01.⁶ Due to this 10-fold error, the court misapplied its own standards in evaluating the reliability of a particular study of the relationship between thyroiditis and low level radiation. These examples, in isolation, are of minimal concern compared to the structural problems detailed below. Isolated mistakes, however, gain power when they are translated into general rules of application. Courts and commentators, therefore, should vigilantly identify and correct mistakes before they are transformed into general legal principles.

Structural misuse

More disturbing than isolated mistakes is the misuse of epidemiology that is certain of repetition because of its incorporation into general legal rules for evaluation of scientific evidence. Some of these rules have been imposed in response to the Supreme Court's decision in *Daubert v. Merrell Dow Pharmaceuticals* (1993), interpreting Federal Rule of Evidence 702 and abandoning the longstanding *Frye* test of admissibility of expert testimony. While the *Frye* test focused on whether an expert's views were "generally accepted" in the scientific community, *Daubert* ushered in a non-exclusive multi-factored inquiry into the expert's methodology that was thought, at the time, to be more flexible in application. Many federal courts, however, have applied *Daubert* to restrict scientific testimony in a way that profoundly departs from scientific principles.

This article reviews six specific types of structural errors made by courts when interpreting epidemiological evidence: (1) the insistence by some courts that "positive human epidemiological studies are always required" to support an opinion that a substance causes a particular adverse health outcome; (2) the conclusion by some courts that the absence of any "positive" epidemiological study trumps

1. *Hall v. Baxter Healthcare Corp.*, 947 F. Supp. 1387 (D. Or. 1996) (excluding testimony of Dr. Shanna Swann); *Muzzey v. Kerr-McGee Chemical Corp.*, 921 F. Supp. 511 (N.D. Ill. 1996) (excluding testimony of New York University Medical School Professor Harry Demopoulos); *Wade-Greaux v. Whitehall Laboratories, Inc.*, 874 F. Supp. 1441, 1482 (D. V.I.) (excluding testimony of Dr. Stuart Newman, a "highly regarded bench scientist"), *aff'd* 46 F.3d 1120 (3d Cir. 1994).

2. *E.g. Berry v. CSX Transportation, Inc.*, 709 So. 2d 552, 559 (Fla. Dist. Ct. App. 1998).

3. Hill, *The Environment and Disease: Association or Causation?* 58 PROC. R. SOC. MED. 295 (1965).

4. Bailey, Gordis, & Greene, *Reference Guide on Epidemiology*, in REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 161 (Washington, DC: Federal Judicial Center, 1994). Although the Federal Judicial Center's second edition of the *Reference Manual on Scientific Evidence* corrected this error, some courts have repeated the *Berry* court's mistake. See *Miller v. Pfizer, Inc.*, No. 99 Civ. 2326, 2002 WL 221410, at *4 (D. Kan. Feb. 7, 2002).

5. *Brock*, 874 F.2d at 312.

6. *In re Hanford Nuclear Reservation Litigation*, No. CY-91-3015-AAM, 1998 WL 775340, at *23 (E.D. Wash. Aug. 21, 1998), *rev'd on other grounds*, 292 F.3d 1124 (9th Cir. 2002).

other nonepidemiological evidence supporting causation; (3) the decision by some courts to prohibit expert testimony in support of causation unless it is based on epidemiological studies that report a doubling of risk associated with exposure; (4) the insistence by some courts that an expert's views can't be considered unless they have been published in peer review publications; (5) the requirement that expert testimony be based only on studies that meet an arbitrary test of "statistical significance"; and (6) the exclusion of expert testimony based on studies with a confidence interval that includes 1.0.

Some courts exclude expert testimony if it is not based on "positive" epidemiological studies. One of the most restrictive and scientifically inadequate rules adopted by some courts is that an expert must rely on epidemiological studies in order to come to a "scientifically reliable" conclusion that a particular exposure caused a particular outcome. For example, some courts have stated arbitrarily that "disinterested and impartial experts in teratology" require "two high quality epidemiological studies" to conclude that a substance is a teratogen (that is, that it causes birth abnormalities).⁷ Some courts have limited this rule to cases involving Bendectin, in which the

mass of epidemiologic evidence reflects no association between the drug and the outcome in question.⁸

The Fifth Circuit in *Brock*, for instance, noted that because there is no consensus that Bendectin is teratogenic, the "most useful and conclusive type of evidence in a case such as this" rests on epidemiology.⁹ According to the court, the lack of "statistically significant" epidemiological proof that Bendectin causes limb reduction defects was "fatal" to the plaintiffs' case.¹⁰ While the court was careful to say that such proof was not necessary in all toxic tort cases, the import of the decision is that epidemiology is necessary when the only other evidence takes the form of animal studies. For, according to the appellate court, a scientist who testifies to an opinion based on in vitro and in vivo animal studies unconfirmed by epidemiology engages in "speculation." Thus, the court held that no reasonable jury, based on such evidence, could conclude that Bendectin caused limb reduction defects.

Arguably, the Bendectin cases occupy a special place in causation jurisprudence because numerous epidemiological studies have failed to uncover an association between the drug and birth defects. For instance, in a Bendectin case, while the U.S. Court of Appeals for the District of Columbia made the broad assertion that the "only way" to extrapolate data from animals to humans was to conduct human experiments or to use epidemiology, it noted that this was especially true where sound epidemiological studies support a conclusion opposite that of nonepidemiological ones.¹¹

The D. C. Circuit recently has emphasized that the rules it developed in the Bendectin litigation can be limited to those cases and others in which there is overwhelming epidemiologic evidence against finding causation.¹² This position is echoed by the Eleventh Circuit.¹³ The rules announced in the Bendectin litigation have, however, been influential in other toxic tort cases.¹⁴ This is true even in cases where, unlike Ben-

dedtin, there is simply a lack of *any* epidemiologic evidence.

For example, the categorical rule from *Brock* was applied in a suit seeking damages for alleged neurotoxic effects of accutane to support the proposition that, without some epidemiological study or statistical basis, an expert's opinion on causation is simply conjecture.¹⁵ And a well-respected federal district court, commenting on evidence presented in the Agent Orange litigation, described epidemiological studies as "the only useful studies having any bearing on causation."¹⁶

As if to underscore the point, a district court judge in Colorado extended the requirement for epidemiology one step further, holding that, in mass exposure cases, the plaintiffs were legally *required* to submit epidemiologic evidence.¹⁷ This was the judge's position despite the fact that the plaintiffs' experts had testified that the exposed community in this case (where plaintiffs alleged contamination of their water supply by hydrazines, trichloroethene, and n-nitrosodimethylamine) was too small to perform an epidemiological study. On appeal, the Tenth Circuit affirmed the district court's decision to grant summary judgment for the defendant, but balked at adopting the lower court's "dicta" that a supporting epidemiological study was required for any mass exposure case.¹⁸

This judicial emphasis on the essential role of epidemiology in establishing medical causation is not reflected in traditional scientific practice. While epidemiology is recognized as a powerful and useful tool in assessing etiologic relationships, many causal associations have been established in the absence of epidemiological proof. In some of these cases, the outcome may be considered a "signature" of the exposure, and pathologic studies, case reports, and animal studies were sufficient to convince the medical community of a causal relationship (*e.g.*, asbestosis and mesothelioma).

Sometimes there is no "signature disease," but scientific evidence aside

7. *DePyper v. Navarro*, No. 83-303467-NM, 1995 WL 788828, at *30 (Mich. Cir. Ct. Nov. 27, 1995).

8. *Ealy v. Richardson-Merrell, Inc.*, 897 F.2d 1159 (D.C. Cir. 1990).

9. *Brock*, 874 F.2d at 311.

10. *Id.* at 313.

11. *Raynor v. Merrell Pharmaceuticals, Inc.*, 104 F.3d 1371, 1375 (D.C. Cir. 1997).

12. *Meister v. Medical Engineering Corp.*, 267 F.3d 1123, 1132 (D.C. Cir. 2001); *Ambrosini v. Labarague*, 101 F.3d 129, 138-139 (D.C. Cir. 1996).

13. *Allison v. McGhan Medical Corp.*, 184 F.3d 1300 (11th Cir. 1999).

14. *Allen v. Pennsylvania Engineering Corp.*, 102 F.3d 194 (5th Cir. 1996); *E.I. Du Pont De Nemours & Co. v. Castillo*, 748 So. 2d 1108 (Fla. Dist. Ct. App. 2000); *Minnesota Mining and Manufacturing Co. v. Atterbury*, 978 S.W.2d 183 (Tex. Ct. App. 1998); *Nelson v. American Sterilizer Company*, 566 N.W.2d 671 (Mich. Ct. App. 1997).

15. *Thomas v. Hoffman-LaRoche, Inc.*, 731 F. Supp. 224 (N.D. Miss. 1989).

16. *In re "Agent Orange" Prod. Liab. Litig.*, 611 F. Supp. 1223, 1231 (E.D.N.Y. 1985) (Weinstein, D.J.).

17. *Renaud v. Martin Marietta Corp.*, 749 F. Supp. 1545, 1554 (D. Colo. 1990).

18. *Renaud v. Martin Marietta Corp.*, 972 F.2d 304, 308 (10th Cir. 1992).

from epidemiology is sufficient to convince physicians that a causal relationship exists. For example, asbestos' relationship with lung cancer was first noted by leading pathologists, to be supported by epidemiologists about a decade later.¹⁹ And the teratogenic effect of thalidomide was discovered through observant clinicians, not through epidemiology.²⁰ Indeed, most teratogens "were initially identified in case reports and clinical studies[,] . . . because teratogenic exposures typically produce qualitatively distinct patterns of congenital anomalies in affected children."²¹

Thus, when courts insist that "reliable" teratologists will not conclude that a particular substance causes birth defects without support from epidemiology, they speak against the weight of history. It is one thing for courts to recognize the insight that epidemiology offers scientists in assessing causation; it is quite another to impose criteria that are not followed by the medical and scientific community, thereby giving epidemiology greater weight in assessing legal causation than it is given by scientists.

Some courts insist that nonepidemiological evidence should play little or no role in an expert's opinion regarding causation. A corollary of requiring epidemiological studies to prove causation is the judicially-invented rule that, where epidemiological studies are inconclusive, other sources of evidence supporting causation cannot reasonably be relied upon by expert witnesses or jurors to find causation. Thus, in the presence of inconclusive or nonexistent epidemiological studies, some courts refuse to allow an expert to testify solely on the basis of animal studies,²² and some judges rule that no reasonable juror could believe that substance X caused outcome Y in an individual, where the expert opinion in support of causation is based on studies other than epidemiology.

Some courts may discount animal studies because the dosage levels are not analogous to human doses.²³ Others specifically note that expert evi-

dence cannot prove causation when relying on chemical structure activity analysis, in vitro studies, and in vivo studies, in the face of an "overwhelming body of contradictory epidemiological evidence."²⁴ Courts will fault scientists for relying on such evidence even where a substance has been classified as a carcinogen by regulatory agencies.²⁵

The U.S. Court of Appeals for the First Circuit limited non-epidemiologic evidence more expansively when it stated that studies of analogous chemical structures, as well as in vivo and in vitro animal studies, "do not have the capability of proving causation in human beings in the absence of any confirmatory epidemiological data."²⁶ Thus, according to the First Circuit, even where there is no epidemiologic evidence at all, other sources of data traditionally relied upon by scientists are not reliable. This position was echoed by the Supreme Court of Pennsylvania when it addressed the scientific methodology of expert witnesses in Bendectin litigation.²⁷ The Pennsylvania court, accepting the defendant's argument, held that the methodology used to assess teratogenicity must rely on epidemiology demonstrating a strong association, while animal studies and chemical analyses could only confirm, not prove, a causal association. Other courts have taken similar positions.²⁸

While most of the vitriol directed towards non-epidemiological studies has been reserved for animal studies, clinically-based case studies also have been identified as particularly unreliable by courts. In a case involving silicone breast implants, a district court stated that "case reports and case studies are universally regarded as an insufficient scientific basis for a conclusion regarding causation because case reports lack controls. . . . Therefore, these cannot be the basis of an opinion based on scientific knowledge under *Daubert*."²⁹ Many other courts have taken similar positions.³⁰

Statements like this take no account of how there was any understanding of cause and effect prior to the first large scale epidemiological studies of the 1950s. As we have

noted, for substances such as asbestos or thalidomide, causal associations were made in the absence of epidemiological studies. All areas of scientific discipline may be relevant to etiologic conclusions: clinical observations, animal studies, toxicologic studies, and chemical analysis. In some cases, epidemiology is the most useful tool for evaluating cause-effect relationships, but not in every case. Some courts' insistence that epidemiological studies trump other forms of scientific evidence is simply contrary to scientific practice.

Some courts prohibit expert testimony in support of causation unless it is based upon epidemiological data that show a doubling of risk associated with exposure to the alleged cause of disease. The two misuses of epidemiology discussed above involve courts excluding evidence that scientists consider relevant to evaluating causal relationships by giving *more* weight to epidemiological

19. Hueper, *OCCUPATIONAL TUMORS AND ALLIED DISEASES* (Springfield, IL: C. C. Thomas, 1942); Hueper, *Cancer in its Relation to Occupation and Environment*, 25 BULL. AM. SOC. CONTROL CANCER 63-69 (1943); Doll, *Mortality from Lung Cancer in Asbestos Workers*, 2 BR. J. IND. MED. 81-86 (1955).

20. Sherman & Strauss, *Thalidomide: a Twenty-five Year Perspective*, 41 FOOD DRUG COSM. L. J. 458-466 (1986).

21. Friedman & Politka, *TERATOGENIC EFFECTS OF DRUGS: A RESOURCE FOR CLINICIANS* vii (Baltimore, MD: Johns Hopkins University Press, 2000).

22. See generally Berger, *Upsetting the Balance Between Adverse Interests: The Impact of the Supreme Court's Trilogy on Expert Testimony in Toxic Tort Litigation*, 64 LAW & CONTEMP. PROBS. 289, 303-304 (2001) (discussing problem with rejecting animal studies and over relying on epidemiological studies).

23. *Merrell Dow Pharmaceuticals v. Havner*, 953 S.W.2d 706, 729 (Tex. 1997).

24. *Richardson by Richardson v. Richardson-Merrell, Inc.*, 857 F.2d 823, 830 (D.C. Cir. 1988).

25. *Allen*, 102 F.3d at 198.

26. *Lynch v. Merrell-National Laboratories*, 830 F.2d 1190, 1194 (1st Cir. 1987).

27. *Blum v. Merrell Dow Pharmaceuticals, Inc.*, 705 A.2d 1314 (Pa. 1997).

28. *Siharath v. Sandoz Pharmaceuticals Corp.*, 131 F. Supp. 2d 1347, 1366-1367 (N.D. Ga. 2001); *Castillo*, 748 So.2d at 1120; *Nelson*, 566 N.W.2d at 674-675.

29. *Wade-Greaux*, 874 F. Supp. at 1411.

30. *Siharath*, 131 F. Supp. 2d at 1361-1362; *Nelson v. American Home Products Corp.*, 92 F. Supp. 2d 954, 969 (W.D. Mo. 2000); *Muzzey*, 921 F. Supp. at 519; *Casey v. Ohio Medical Products*, 877 F. Supp. 1380, 1385 (N.D. Cal. 1995); *Atterbury*, 978 S.W.2d at 199. But see, e.g., *Silivanch v. Celebrity Cruises, Inc.*, 171 F. Supp. 2d 241, 264 (S.D.N.Y. 2001) (case reports "are plainly relevant to an expert's opinion as to whether a given risk factor is generally associated with an injury").

studies then do scientists. There is, however, another variation on the theme. Courts also exclude epidemiologic evidence that scientists value by giving some epidemiological studies less weight than do scientists.

Thus, some courts insist that, to be heard by a jury, expert opinions must be based on epidemiological studies that meet artificial, non-scientific standards of "scientific validity." Perhaps the most dangerous of these court-created criteria is the requirement that any epidemiological study relied on to support causation demonstrate an association between an exposure and disease of more than twice the background incidence of disease. In essence, these courts have conflated the magnitude or strength of association revealed in a *population-based* study with the probability that a substance has caused disease in a particular *individual*.³¹ By imposing this requirement, courts appear to have been directly influenced by scientists who resurrected the concept of "probability of association" in an attempt to rationalize compensation for radiation-induced cancers.³² The derivation of individual "probabilities" of causation from population-based data has been

roundly criticized by epidemiologists.³³ Yet it retains force in the courts, and informs the rule adopted by some judges that expert testimony is inadmissible unless it relies on an epidemiological study with a relative risk (or odds ratio) of 2.0 or more.

Examples of this misapplication of probability-based concepts abound in toxic tort cases, from those involving intrauterine devices, per chloroethylene exposure, to silicone breast implants. One court, along with imposing this arbitrary requirement on a testifying expert, also rejected the expert's reliance on a study that did find a relative risk higher than 2.0, because in the court's view, a larger sample size was required than the 445 women studied.³⁴

In 1995, a panel of the Ninth Circuit went perhaps the farthest of courts in imposing unscientific standards on expert evidence, holding not just that a witness must rely on an epidemiological study with a relative risk greater than two, but that a study showing less than a relative risk of two "may suggest teratogenicity" but "actually tends to disprove legal causation."³⁵ Building on the Ninth Circuit's statements, a federal district court judge held that plaintiffs had to

submit epidemiological studies showing a two-fold increase in risk to proceed in the massive Hanford Nuclear Reservation Litigation.³⁶ In 2002, the Ninth Circuit reversed the district court's decision, cabining its previous 1995 decision to circumstances where there is no other evidence to support causation other than epidemiology. Between 1995 and 2002, however, all district courts in the Ninth Circuit were required to follow the 1995 decision—and making matters worse, many district courts outside of the Ninth Circuit's jurisdiction relied on that court's 1995 decision to impose their own similar requirements.³⁷

In theory, the requirement that any epidemiological study presented to a jury report a two-fold magnitude of association might make sense if the plaintiffs rely solely on epidemiology to prove causation. In such a case, there would not be sufficient evidence to conclude that the specified exposure more likely than not caused the disease in the plaintiff. If, however, as is almost always the case, the plaintiffs also introduce other evidence to prove causation—such as pathology, animal experimentation, molecular modeling, or case studies—then requiring the epidemiological studies relied on by the plaintiffs to show a relative risk or odds ratio of greater than 2.0 is without support. An appeals court in New Jersey got it right when it stated that this requirement "makes little sense, scientifically or legally."³⁸

There are also more subtle problems with the thinking evinced by the courts above, all of which have been astutely observed by epidemiologists. To begin with, some studies may only measure an increased incidence in the subset of cases which would not have occurred had there been no exposure (excess cases) and may ignore other cases in which exposure played a role in the etiology of the disease (etiologic cases).³⁹

For instance, if an individual develops a disease five years earlier than she would have had she not been exposed to a certain substance, she would not be considered an excess

31. E.g., *Magistini v. One Hour Martinizing Dry Cleaning*, 180 F. Supp. 2d 584, 591 (D.N.J. 2002) (citing Federal Judicial Center's 2000 *Reference Guide on Epidemiology* for proposition that "the threshold for concluding that an agent was more likely than not the cause of an individual's disease is a relative risk greater than 2.0") (emphasis in original) (internal quotation marks omitted); but see e.g., *Miller*, 2002 WL 221410, at *12 (rejecting as "arbitrary" the requirement that relative risk be greater than 2.0).

32. Council on Scientific Affairs, *Radioepidemiological Tables*, 257 J.A.M.A. 806–809 (1987).

33. Greenland & Robins, *Conceptual Problems in the Definition and Interpretation of Attributable Fractions*, 128 Am. J. Epi. 1185–1197 (1988).

34. *Kelley v. American Heyer-Schulte Corp.*, 957 F. Supp. 873, 878, 880 (W.D. Tex. 1997).

35. *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 43 F.3d 1311, 1321 (9th Cir. 1995).

36. *In re Hanford Nuclear Reservation Litigation*, 1998 WL 775340, at *13, *rev'd*, 292 F.3d 1124 (9th Cir. 2002).

37. *Pozefsky v. Baxter Healthcare Corp.*, No. 92 Civ. 0314, 2001 WL 967608, *3 (N.D.N.Y. Aug. 16, 2001); *Siharath*, 131 F. Supp. 2d at 1356; *In re Breast Implant Litigation*, 11 F. Supp. 2d 1217, 1225–1227 (D. Colo. 1998).

38. *Grassis v. Johns-Manville Corp.*, 591 A.2d 671, 676 (N.J. App. Div. 1991).

39. Greenland and Robins, *supra* n. 33, at 1185. "Relative risk" is calculated by dividing the incidence rate of an illness or outcome in an exposed

population by the incidence rate of that illness or outcome in a general or presumably unexposed population. For example, if an illness occurs in 5 out of 1,000 people in the general population but 20 out of 1,000 in the exposed population, the relative risk is $.02/.005 = 4.0$. A relative risk of 2 would reflect an exposed incidence rate twice as large as a baseline incidence rate. Closely related to relative risk is the "attributable risk," which provides an estimate of the "excess cases" attributable to exposure. Attributable risk is calculated by taking the difference between the incidence rate in the exposed group and the incidence rate in an unexposed but presumably similar group and dividing that difference by the incidence rate in the exposed group. In the example given above, the attributable risk would be $(.02-.005)/.02 = .75$. That figure implies that three-quarters of the cases in the exposed population are statistically attributable to exposure. These are the "excess cases," in excess of the number that would be expected to occur in the absence of exposure to the putative cause. But it is crucial to keep in mind two matters: First, these probability analyses operate at the population level and do not automatically apply to the individual case; and second, even at the population level, these kinds of calculations presume perfect epidemiological studies, unrealistically free of bias or error and with perfect matching of the exposed and unexposed groups on all factors except exposure. Judgment is always necessary in assigning weight to epidemiological evidence, whether in science or in law.

case, but she would be considered an etiologic case. From the law's perspective (and science's) the fraction of all etiologic cases attributable to exposure is significant, not just the fraction of excess cases associated with exposure. This is because developing a serious disease at age 45 instead of age 50 will likely have an impact on overall life expectancy. If a court imposes a requirement that an injured plaintiff may only recover if she can prove that the probability that her disease was induced by exposure exceeds 50 percent, use of an excess fraction (a fraction that only represents the percentage of excess cases attributable to exposure) would disadvantage the plaintiff.

Courts often demonstrate their failure to understand the significance of etiologic cases by focusing only on excess cases. For instance, the Supreme Court of Texas, in *Merrell Dow Pharmaceuticals v. Havner* (1992), stated that "there is a rational basis for relating the requirement that there be more than a 'doubling of the risk' to . . . the more likely than not burden of proof."⁴⁰ The court hypothesized a population in which a condition "naturally" occurs in six out of every 1,000 people. Then the court imagined that, of 1,000 people taking a drug, nine contracted the disease. While acknowledging that the model is an "oversimplification," the court baldly stated that it is not "more likely than not" that the drug caused any one incidence of disease. The *Havner* court did not even consider the possibility that, while there were only three excess cases in the population exposed to the hypothetical drug, the other six cases might have developed disease five years earlier because of exposure to the drug. The fact that an exposure causes susceptible individuals to develop disease earlier than they would have absent exposure is both legally and medically significant.

Moreover, courts that insist on a two-fold magnitude of association between determinant and disease are ignoring the difficulty of estimating individual risk from population-based data. While population-based data

are useful in evaluating issues of general causation in a toxic tort case, they must be carefully applied to the question of specific causation. The basic premise of the probability-of-causation approach—that the population-based data are precisely replicated in each individual—"holds only if the individual is truly representative of the reference population."⁴¹ The analysis assumes, in the absence of confounding, selection bias, or misclassification, that the background rate of disease is the same for all non-exposed cohort members.⁴² Given unmeasured genetic and environmental factors, this assumption is likely false.⁴³ The court-created rule that any expert testimony offered in support of causation be supported by an epidemiological study with a greater than two-fold magnitude of association ignores this reality.

The reasoning employed by courts in arriving at the two-fold risk requirement also ignores an important complexity of disease process. Individuals vary in their response to a given disease determinant, depending on many factors. A substance might, for example, have different effects depending on the age of the exposed individual. Depending on how data are stratified, a study that reveals no statistical association between the exposure and the disease might misleadingly suggest that there is no association, whereas it is the complexity of the association that is being misassessed.

Relying solely on the magnitude of association to measure the actual strength of the studied biological relationship overlooks the fact that the strength of a factor's effect on a population depends on the relative prevalence of its causal complements.⁴⁴ Most diseases are thought to be caused by multiple unrelated factors (or component causes), making up, for lack of a better analogy, a "causal pie." Additionally, there might be many different causal pies for one disease, each of which, when all the component causes are present, is sufficient to lead to disease.

Whatever the biologic significance of a particular "cause" of disease, if

the other component causes that make up the same sufficient causal pie are rare, then the magnitude of association for the particular "cause" will be small. Given this, it is possible that over a span of time, the magnitude of association between a particular cause and a particular disease "may change because the prevalence of its causal complements in various mechanisms may also change."⁴⁵ There are several examples of accepted causal relationships that have a relatively weak strength of association, such as cigarette smoking and cardiovascular disease or passive smoking and lung cancer, demonstrating the relative value that strength of association plays in the scientific world.⁴⁶ Some courts, however, incorrectly assume that a larger magnitude of association always indicates a greater likelihood of causal relationship—and vice versa. This is a drastic and erroneous oversimplification.

Some courts will not admit expert testimony if the expert has not published her views in a peer-reviewed scientific journal. Another prevalent legal misuse of epidemiology is the insistence by some courts that experts cannot testify if their opinions—or the epidemiological analyses on which they are based—have not been subjected to the peer-review process. Peer review and publication have even been called the "most important means of ensuring that an expert's methodology is sound."⁴⁷

While on its face the peer-review requirement may seem reasonable, it is not reflective of the scientific process of decision making. Publication in a peer-reviewed journal may ensure that a particular study or methodology receives attention, but

40. 953 S.W.2d 706, 717 (Tex. 1997).

41. Council on Scientific Affairs, *supra* n. 32, at 807.

42. Robins & Greenland, *Estimability and Estimation of Expected Years of Life Lost Due to a Hazardous Exposure*, 10 STAT. MED. 79-93, at 80 (1991).

43. *Id.* at 81.

44. Rothman & Greenland, *MODERN EPIDEMIOLOGY* 10-11 (Philadelphia, PA: Lippincott-Raven, 1998).

45. *Id.* at 11.

46. *Id.* at 24.

47. *Jones v. U.S.*, 933 F. Supp. 894, 897 (N.D. Cal. 1996); see also *Havner*, 953 S.W.2d at 727.

it is no guarantee that the study is particularly reliable. Nor is the fact that an opinion has not been published in a peer-reviewed journal evidence that the opinion lacks scientific reliability. This is reflected in the Supreme Court's opinion in *Daubert*, which noted the fact that some well-grounded theories will not be published, and some theories or techniques will be too new or of too limited interest to be published.⁴⁸

In addition, scientists often disagree on the interpretation of each other's data. One of the primary purposes for publication is to generate debate and discussion in the scientific community. Not all of this debate will take place in the pages of peer-reviewed journals. Epidemiologists do not refuse to consider other epidemiologists' views simply because they are not published in peer-reviewed journals. Nor should the courts. To insist on peer-reviewed publication of an expert's views before allowing a jury to consider the expert's testimony is inappropriate both because the "actual practice of medicine" does not require it and because "victims of a new toxic tort should not be barred from having their day in court simply because the medical literature . . . has not yet been completed."⁴⁹

Some courts require that expert testimony be based only on studies with results that meet an arbitrary test of "statistical significance." Another way in which courts misuse

epidemiology is in their treatment of hypothesis testing, which is one way for epidemiologists to estimate the accuracy of an association between exposure and disease revealed in a study. Epidemiologists do so by measuring how likely it is that the same result would have obtained had there been no true association (*i.e.*, if the null hypothesis were true).

The "p-value" is the numerical expression of this likelihood. Many courts require that a study's results have a p-value of less than 0.05 (*i.e.*, that there is less than a 5 percent likelihood that the study's results would have obtained had the null hypothesis been true) before allowing an expert to rely on the study for an admissible opinion. Misusing the language of science, these courts say that, because this number must be met to ensure "statistical significance," studies that fail to meet it simply cannot be considered. One district court described a p-value of less than 0.05 as "[t]he most common value used to establish significance and to say that an observed association is probably real."⁵⁰

As epidemiologists know, however, 0.05 is an arbitrary number, with more historical than inferential value.⁵¹ The division of study results into "significant" and "non-significant" serves no purpose in causal investigation other than to mislead. Rather, use of bright-line rules in the context of current scientific and judicial analysis "stems from the apparent objectivity and definitiveness of the pronouncement of significance" and "can serve as a mechanical substitute for thought, promulgated by the inertia of training and common practice."⁵² The selection of a certain level of "significance" at which to test a hypothesis involves a balance between the number of false positives and false negatives considered acceptable. Using a higher p-value (say, 0.10) will increase the number of false positives, but decrease the number of false negatives (although there is not a one-to-one relationship between the two). There is no absolutely correct p-value from which to choose, but judges are prone to

use 0.05 simply because they are told that this is the standard for "statistical significance." Indeed, some judges, with the knowledge that overreliance upon tests of statistical significance has been criticized by scientists themselves, insist that "[t]here must be some objective way to put a value on what the study says or shows."⁵³ If scientists and courts insist on "significance" testing, the actual p-value should be reported and considered, not simply whether it falls above or below an arbitrary point.

Some courts misunderstand the utility of a study's confidence interval and exclude expert testimony based on studies with an interval that includes 1.0. Finally, some courts state as a matter of law that if a confidence interval (usually performed at the 95 percent level) includes the value 1.0, then the study is not "statistically significant," and therefore is not reliable. Many courts that take this position cite the Fifth Circuit's decision in *Brock* as support.⁵⁴ The imposition of this legal requirement flows from a misunderstanding of the basic reason for using confidence intervals. The *Havner* court stated that a confidence interval "tells us if the results of a given study are statistically significant at a particular confidence level."⁵⁵ According to the court, if the confidence interval includes the number 1.0, then it is not statistically significant, and hence inconclusive.

This equivalence of the confidence interval with hypothesis testing is precisely contrary to the purpose of providing a confidence interval. The confidence interval is intended to provide a range of values within which, at a specified level of certainty, the magnitude of association lies.⁵⁶ The confidence interval is not another way to conduct hypothesis testing. This point was brought home by a reanalysis of 71 clinical trials that relied on hypothesis testing to conclude that there was no relationship between a proposed treatment and a disease response, while the use of confidence intervals revealed a moderate to strong effect of the treatment being tested.⁵⁷

48. *Daubert*, 509 U.S. at 593.

49. *Turner v. Iowa Fire Equip. Co.*, 229 F.3d 1202, 1209 (8th Cir. 2000); *Heller v. Shaw Indus., Inc.*, 167 F.3d 146, 155 (3d Cir. 1999).

50. *Wade-Greaux*, 874 F. Supp. at 1452.

51. Rothman & Greenland, *supra* n. 44.

52. *Id.* at 187.

53. *Magistrini*, 180 F. Supp. 2d at 605 n. 25 (stating that "even if some sciences don't require [a specific p-value] this Court does").

54. *Grant v. Bristol-Meyers Squibb*, 97 F. Supp. 2d 986, 992 (D. Ariz. 2000); *Flue-Cured Tobacco Cooperative Stabilization Corp. v. EPA*, 4 F. Supp. 2d 435, 461 (M.D. N.C. 1998); *LeBlanc v. Merrell Dow Pharmaceuticals, Inc.*, 932 F. Supp. 782, 783 (E.D. La. 1996).

55. *Havner*, 953 S.W.2d at 723.

56. Rothman & Greenland, *supra* n. 44, at 190.

57. Freiman, et al., *The Importance of Beta, the Type II Error and Sample Size in the Design and Interpretation of the Randomized Control Trial: Survey of 71 "Negative" Trials*. 299 N. ENGL. J. MED. 690-694 (1978).

A study in which the 95 percent confidence interval includes the value 1.0 (*i.e.*, a relative risk in which the incidence of disease is no greater in those exposed to the putative cause than in those unexposed) may be statistically compatible with no association, but could be overwhelmingly compatible with a strong association. Imagine two different studies. One finds a relative risk of 1.8, with a 95 percent confidence interval between 1.05 and 3.08. Another finds a relative risk of 3.4, with a 95 percent confidence interval between 0.95 and 12.17. It is true that, of these two studies, it is less likely that the magnitude of association observed in the first would have been observed had there been no causal relationship between the disease and the proposed causal factor. This does not mean, however, that the observations of the second study (or, for that matter, the first) offer no support for the proposition that there is a causal relationship present. Simply put, when judges only use confidence intervals to determine whether 1.0 lies inside or outside the interval, they are acting contrary to the scientific process by abandoning the inferential value of the confidence interval.

Results of misuse

Theoretically, one might look at the examples of misuse described above and conclude that, while unfortunate, disparate treatment of epidemiology in the scientific and legal arenas should not raise the hackles of judges, lawyers, scientists, or policy makers. In reality, however, the barriers that judges have created to limit the admissibility of expert testimony have profound, one-sided consequences. In the short-term, the structural limitations on admissible evidence mostly function to disfavor plaintiffs seeking compensation for injuries and assist companies seeking to avoid liability (although there will be some instances in which corporate defendants will be disadvantaged by these rules). In the long run, they ensure that our system of justice will be increasingly divorced from the truth—and that there will be a declin-

ing respect for both science and law.

The short-term harm to plaintiffs is due to several factors. First, plaintiffs carry the burden of proof in civil cases, and must show that a defendant's product or conduct caused their injury by a preponderance of the evidence. The rules described above make proving causation more difficult. This effect might be mitigated somewhat if the rules were accompanied by their mirror images. In other words, if an epidemiological study reporting a relative risk of greater than 2.0 was not only necessary to support expert testimony on causation, but sufficient to prove causation, then the rule, though still irrational, would have a less inequitable effect. We do not suggest, however, that courts adopt such a rule.

By requiring the presentation of epidemiological data to support testimony on causation, courts also effectively disfavor injured plaintiffs because of unequal access to resources. Injured plaintiffs normally do not have sufficient funds to support epidemiologic research related to their legal claims, and federal funding of large-scale epidemiological studies has decreased along with funding of scientific research in general. In contrast, most defendants in mass toxic tort cases have sufficient funds to choose which relationships to study, how to study them, and whether to publicize the results. Some courts have recognized this problem.⁵⁸

Finally, there are short-term effects on injured plaintiffs because of the varying interests of repeat and non-repeat players in the justice system. Most plaintiffs have no interest in funding a long-term epidemiological study that may have no impact on their particular case, or may take too long to have an impact. Corporations, with the expectation of being sued multiple times, have a greater incentive to fund epidemiological studies, and to publish those that are favorable to their defense.

The long-term effect of the misguided restrictions on evidence described in this article is simple. As more and more relevant evidence is

excluded, the outcomes in court cases will become less and less consistent with the truth. This, we fear, will concomitantly lead to decreased respect by the public for science and for the law.

Causes of legal misuse

In our view, there are several reasons why the law continues to misuse epidemiology. First, some of the misuse stems from the unfamiliarity of many judges with the scientific process. Miscommunication and misunderstanding inevitably lead to misuse.

A second explanation is the desire of judges to impose bright-line rules, such as requiring an epidemiological study to demonstrate a two-fold magnitude of association, a p-value of less than 0.05, or a confidence interval that does not include unity. Judges, seeking certainty where scientists are unwilling to impose it, may apply these rules to make decisions easier, and give an air of scientific "objectivity" to their rulings.

Third, there may be a fear (explicitly articulated in some of the *Bendectin* cases) that "junk science," in combination with dramatic testimony from injured plaintiffs, will work a pernicious influence on gullible juries. If this fear is truly justified, however, then the jury system should be improved. The evidentiary rules should not be rigged so that scientific studies are given more or less weight in the courtroom than scientists would give them, depending upon what is necessary to keep the case from the jury.

A fourth factor contributing to legal misuse of epidemiology is a combination of judicial hubris and susceptibility. Judges are called upon, every day, to master unfamiliar subjects and to make major decisions affecting others' lives. They do so, moreover, in an adversarial context, where they are understandably suspicious of all of the participants' motives. Ultimately, even without any scientific training, some judges come

58. *Bonner v. ISP Technologies, Inc.*, 259 F.3d 924, 928-930 (8th Cir. 2001); *Marder*, 630 F. Supp. at 1094.

to believe that they are better equipped than scientists (or at least the scientists before them) to determine what is and is not valid scientific methodology.

An even more basic reason these rules have been adopted is that their substantive results are desired by those who benefit from them the most—potential corporate defendants seeking to minimize their liability and maximize their profits. In saying this, we are not ascribing any pernicious motives to these companies. They are doing exactly what one would expect profit-maximizing institutions in our economy to do—trying to get the legal rules (and judges' interpretations of them) changed to their benefit. While demonstrating this fact is well beyond the scope of this article, we will provide one anecdote that reflects the means by which some companies, and their supporters, further their goals.

Recently, America's largest corporations helped launch a massive campaign to convince policy makers that "junk science" was an enormous threat to our system of justice—and that the legal rules had to be changed to prevent it. The centerpiece of the campaign was a book written by Peter Huber, detailing fantastical claims of the willingness of judges and juries to rely on the "far fringes of science" to assess liability.⁵⁹ Huber described juries (and some judges) running amok, awarding damages to plaintiffs who brought "frivolous" claims based on questionable scientific evidence. The book was cited by several courts, including the Ninth Circuit in its *Daubert* decision, and undoubtedly played a key role in persuading some judges that changes in the legal rules governing the admissibility of scientific evidence were needed. Much of Huber's book, however, was blatantly unreliable.⁶⁰ Huber's ideas appear to

have gained success primarily due to the efforts of the Manhattan Institute for Policy Research, a conservative think tank, to promote the ideas of people like Huber and others who would raise barriers to a plaintiff's ability to successfully bring suit against a corporation.⁶¹

Suggestions for action

We propose several possible ways to improve the law's use of science in evaluating causal relationships. First, to the extent that legal misuse of epidemiology is caused by misunderstanding or miscommunication, both the scientific and legal communities have to do more to educate judges about science. Second, there are procedures for judges to appoint experts for the court, not to take sides, but, rather, to ensure that the judge, and in some cases the jury, understands the relevant scientific principles in a case. The American Association for the Advancement of Sciences has launched a program to help federal judges identify potentially-helpful court-appointed experts—it became fully operational two years ago and since that time has assisted federal judges in identifying experts in fields ranging from engineering to econometrics to epidemiology.⁶²

Third, the scientific community needs to provide feedback to the legal community in general, and judges in particular, about whether science is being used correctly. The simplest way for scientists to do this is to inform judges and litigants directly when they learn of a lawsuit or legal decision in which science is being misused. Scientists also should publish articles, in legal as well as scientific media, expressing their concerns and correcting the errors they have found. In our experience, most judges are surprisingly responsive to this feedback; they want to use science correctly. The most formal, and most effective, means to provide such feedback is for scientists to file *amicus curiae* briefs. *Amicus* briefs, filed either while a decision is under reconsideration or in advance of an appellate argument regarding an important scientific question, are

essential for informing judges how to use scientific principles correctly.

Scientific organizations and educational institutions also should promote educational programs to bring scientists, lawyers, and judges together. Many judges are particularly eager to learn from experts who are not active participants in the adversarial dispute before them. Scientific organizations also should work with legal institutions to establish more formal feedback processes and enhance legal understanding of science. The Administrative Office of the U.S. Courts, the Federal Judicial Center, the National Center for State Courts, and the National Judicial College all should be interested in developing such processes.

Finally, increased funding of scientific research also will help law use science to find the truth. Neither law nor science is likely to discover the truth if scientific research is funded primarily by those who have a strong financial interest in a particular outcome. Admittedly, government itself has an incentive in preserving long-held government positions (such as those related to health effects of low-dose radiation), but increased government funding is part of a means to ensure that science and law both have a better chance to do what we want them to do—help us discover the truth. ☛

ARTHUR H. BRYANT

is the Executive Director of Trial Lawyers for Public Justice, a national public interest law firm with offices in Washington, DC, and Oakland, CA (abryant@tlpj.org).

ALEXANDER A. REINERT

is an Associate at Koob & Magoolaghan, a civil rights law firm located in New York City. (aar@kmlaw-ny.com).

59. Huber, *GALILEO'S REVENGE: JUNK SCIENCE IN THE COURTROOM* (New York, NY: Basic Books, 1991).

60. Chesebro, *Galileo's Retort: Peter Huber's Junk Scholarship*, 42 AM. U. L. REV. 1637-1726 (1993).

61. *Id.* at 1707-1722.

62. Information about the Association's project can be found at <http://www.aaas.org/spp/case/case>.