Epidemiology in the Legal Arena and the Search for Truth

Arthur H. Bryant and Alexander 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 presentation is devoted to exploring and improving the means by which law relies on scientific principles, 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. Over 35 years ago, in Brown v. Board of Education (1, pp. 494–5), the US Supreme Court relied on social science research to reject the notion that racially separate education was really equal. The law's interest in and reliance on science has been growing ever since. 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 who loses) in a wide array of contexts, including toxic tort, employment discrimination, environmental protection, product liability, civil rights, and criminal cases.

In the past 20 years, however, the law has misused one discipline in particular, epidemiology. 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 are establishing systemic rules that give epidemiologic studies either more or less weight than scientists would give them. As a result, the law is being taken further from the truth.

This presentation reviews six ways in which some courts are misusing epidemiologic studies. These errors often preclude the courts and (especially juries) from considering evidence that scientists (and especially epidemiologists) would readily consider. And, perhaps most galling, the justification for precluding reliance upon this evidence is that it is not really "scientific." This presentation offers some explanations for the errors, it contends that epidemiologic studies should be given the same weight and consideration in the legal arena that they are given in the scientific arena, and finally suggests several ways that epidemiologists, lawyers, judges, and other policy-makers can work together toward this goal.

ISOLATED LEGAL MISUSE OF EPIDEMIOLOGY

Jurors and judges in state and federal court increasingly are called upon to evaluate the strength of a litigant's case based in part upon the presentation of epidemiologic studies. Judges initially must decide whether scientific testimony is admissible, and subsequently decide, after trial, whether there is sufficient evidence to support the jury's verdict. In either deliberation, judges must critically consider the nature of the evidence presented, including the role that epidemiology played in informing the scientific testimony. Thus, an initial problem for courts assessing epidemiologic evidence 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, a Florida appellate court's assertion that a "generally accepted set of standards for evaluating epidemiological studies is known as the Koch [sic] Postulates" (2, p. 559). The court went on to list not Koch's Postulates, but a modified version of Hill's suggestions for analyzing epidemiologic studies (3). Fortunately, the court was right on the fundamentals, and its error is understandable in part, given that Federal Judicial Center's 1994 Reference Manual on Scientific Evidence, upon which the court relied, also misidentified Hill's criteria as Koch's Postulates (2).

In some cases, however, the mistakes made by courts in evaluating epidemiologic evidence carry more substantive consequences. For example, the US Court of Appeals for the
Fifth Circuit, in *Brook v. Merrell Dow Pharmaceuticals, Inc.* (4), considering the sufficiency of scientific evidence presented in a Bendectin case, illustrated the difficulty some courts have in 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 incorrect, but *Brook* has not been cited in any subsequent case for this proposition. In addition, the court, without specifying at what level the confidence interval was being performed, stated that if a confidence interval included unity in its range of possible values, then “no statistically significant conclusions could be drawn” from a study (4, p. 312). The treatment of confidence intervals as interchangeable with “significance testing”—a proposition for which *Brook* has been cited by numerous courts—is a common misunderstanding, and will be discussed in detail below in the section on Structural legal misuse of epidemiology.

The US Court of Appeals for the Second Circuit also has demonstrated confusion about epidemiologic concepts. In an asthessos/colost cancer case, for instance, the court correctly chastised a district court for insisting on a standardized mortality ratio of 150 in order to prove causation (5, p. 1134). At the same time, however, the court conflated the concept of magnitude of association with the distinct concept of statistical significance. While the Second Circuit surely was correct to observe that a study could have a standardized mortality ratio of less than 150 and nonetheless be “statistically significant,” its suggestion that there is some connection between the magnitude of association reported by a study and the study’s significance is just as surely a departure from epidemiologic principles.

Similar examples abound. In a lengthy opinion from the Eastern District of Washington, a federal judge starkly demonstrated the difficulty that some courts have with basic statistical method. After noting that accepted *p* values for testing statistical association included “5 percent” and “1 percent,” the court equated these percentage with the numerical values 0.05 and 0.1, respectively (6, p. 22). Due to this 10-fold error, the court misapplied its own standards in evaluating the reliability of a particular study of the relation between thyroiditis and low level radiation.

These examples, in isolation, are of minimal concern compared with 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 LEGAL MISUSE OF EPIDEMIOLOGY**

While isolated mistakes raise some cause for concern, what is more disturbing is the misuse of epidemiology that is certain of repetition because of its incorporation into general legal rules for exclusion and 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* (7), which announced a change in the evidentiary approach to admissibility of scientific testimony. Prior to *Daubert*, most federal appellate courts followed the *Frye* rule of admissibility of expert scientific testimony, taken from a 1923 case in the US Court of Appeals for the DC Circuit, which asked if the expert’s views and methodology were “generally accepted” in the scientific community. *Daubert* interpreted Federal Rule of Evidence 702 (the rule governing admissibility of expert testimony) to abandon the *Frye* test for a non-exclusive multifactor inquiry into the expert’s methodology. The *Frye* test continues to have force in those state courts that have not yet adopted the *Daubert* standard.

A federal district court case, *Wade-Greaux v. Whitehall Laboratories, Inc.* (8), involving the alleged teratogenic effect of an asthma medication illustrates many of the problematic structural rules that have emerged in the wake of *Daubert*. The *Wade-Greaux* court first asserted the primacy of epidemiology in assessing causation, relying on defense testimony to conclude that “positive human epidemiological studies are always required to reach a conclusion as to whether a specific agent is teratogenic in humans” (8, p. 1451). The court described a “positive” epidemiologic study as one that presents a statistically significant (defined as a study with a *p* value of less than 0.05 or a 95 percent confidence interval excluding unity) association between a determinant and disease, and stated that epidemiologists do not draw any conclusions regarding causal association until several positive studies have linked a specific exposure with a particular outcome.

The court then enumerated and applied several factors to conclude that each of the plaintiff’s expert witnesses offered inadmissible testimony (8, pp. 1479–81). First, because the witnesses testified to a causal relation in the absence of positive epidemiologic studies, they were “out of sync” with the teratology community, and, hence, their methodology was too novel to be considered by a jury. Second, the court noted that, because the methodology of the plaintiff’s witnesses had not been accepted in peer review publications, it failed the requirement that there be a “specialized literature” utilizing the methodology of the witness. The court also found that none of the plaintiff’s witnesses put their methodology to use outside of the courtroom because they did not engage in any activities in the field of teratology (one was a clinician, and the other was a pharmacologist and toxicologist). The court faulted the plaintiff’s experts for relying on case reports and animal studies because of the high rate of error in extrapolating such data to humans. The court finally concluded that none of the plaintiff’s witnesses were qualified to testify because the witnesses did not, as part of their regular activity, study the causes of birth defects in humans. Thus, the evidence offered by the witnesses was not admissible for the purposes of proving causation.

Adding insult to injury was the terminology used to justify the exclusion of plaintiff’s proposed expert witness testimony in *Wade-Greaux*. The judge said the jury could not consider the testimony of the plaintiffs’ five proposed expert witnesses because they had not followed proper “scientific methodology” and their views were not “scientific knowledge” (8, p.
1477). This terminology is not unusual. In most of the cases excluding proposed scientific testimony, the court says the testimony cannot be considered because it is not "scientific." On this basis, the views of highly-accomplished scientists have been barred from the courts, including those of Dr. Shanna Swan (9), New York University Medical School Professor Harry Demopoulos (10), and Dr. Stuart A. Newman, a "highly regarded bench scientist" (8, p. 1482). Rhetorical symbolism aside, however, the Wade-Greaux decision is notable because it applies many of the structural misuses of epidemiology in the law that we engage in detail below.

**Epidemiologic studies required for causation**

One of the most restrictive, and scientifically inadequate, rules adopted by some courts, including in Wade-Greaux, is that an expert must rely on epidemiologic studies in order to come to a "scientifically reliable" conclusion that a particular exposure caused a particular outcome. One Michigan state court, while applying the Frye test for admissibility, held that "disinterested and impartial experts in etiology" require "two high quality epidemiological studies" to conclude that a substance is a teratogen (11, p. 30). Some courts have limited this rule to cases involving Bendectin, in which the mass of epidemiologic evidence reflects no association between the purported determinant and the outcome in question (12). The Fifth Circuit's opinion in Brock, for instance, considering the sufficiency of scientific evidence presented in a Bendectin trial (i.e., is it reasonable for a juror, given the evidence presented, to conclude that the drug caused birth defects), 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 (4, p. 311). According to the court, the lack of "statistically significant" epidemiologic proof that Bendectin causes limb reduction defects was "fatal" to the plaintiffs' case (4, p. 313). 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, 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 epidemiologic studies have failed to uncover an association between the drug and birth defects. For instance, while the US Court of Appeals for the District of Columbia Circuit made the broad assertion that the "only way" to extrapolate data from animals to humans was to conduct human experiments or use epidemiology, it noted that this was especially true where sound epidemiologic studies support a conclusion opposite that of non-epidemiologic ones (13, p. 1375). The DC 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 (14, p. 424). This position is echoed by the Eleventh Circuit (15). The rules announced in the Bendectin litigation have, however, been influential in other toxic tort cases (16-19). This is true even in cases where, unlike the Bendectin cases, there is simply a lack of any epidemiologic evidence. For example, Brock was applied in a suit seeking damages for alleged neurotoxic effects of acetanilid to support the proposition that, without some epidemiologic study or statistical basis, an expert's opinion on causation is simply conjecture (20). And federal district court Judge Jack Weinstein, commenting on evidence presented in the Agent Orange litigation, described epidemiologic studies as "the only useful studies having any bearing on causation" (21, p. 1231).

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 (22, p. 1554). This was the judge's position despite the fact that the plaintiffs' experts had testified that the exposed community in the case (where plaintiffs alleged contamination of their water supply by hydrazines, trichlorethene, and n-nitrosodimethylamine) was too small to perform an epidemiologic 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 epidemiologic study was required for any mass exposure case (23, p. 308).

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 relations, many causal associations have been established in the absence of epidemiologic 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 relation (e.g., asbestos with asbestosis and mesothelioma). Sometimes there is no "signature disease," but scientific evidence aside from epidemiology is sufficient to convince physicians that a causal relation exists. For example, asbestos' relation with lung cancer was first noted by leading pathologists, to be supported by epidemiologists about a decade later (24-26). And the teratogenic effect of thalidomide was discovered through observant clinicians, not through epidemiology (27).

Thus, when courts insist, as did the Wade-Greaux court (8, p. 1451), that "reliable" epidemiologists 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.

**Epidemiologic studies trump other evidence**

A corollary of requiring epidemiologic studies to prove causation is the judicially-invented rule that, where epi-
demographic studies are inconclusive, other sources of evidence supporting causation cannot reasonably be relied upon by expert witnesses or jurors to support causation. Thus, in the presence of inconclusive or nonexistent epidemiologic studies, some courts refuse to allow an expert to testify solely on the basis of animal studies (11, p. 32), 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 (28, p. 729). Others specifically note that expert evidence 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" (29, p. 830). For example, one court excluded testimony regarding an association between ethylene dichloride and brain cancer for lack of the expert's reliance on significant epidemiologic studies despite the substance's classification as a carcinogen by regulatory agencies (16, p. 198).

The US Court of Appeals for the First Circuit limited non-epidemiologic evidence more expansively when it stated that study 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" (30, p. 1194). 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 (31). The Pennsylvania court, accepting Merril Dow'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, an epidemiologic association. Other courts have taken similar positions (17, p. 1120; 19, pp. 674–5; 32, pp. 1366–7).

While most of the vitriol directed towards studies other than epidemiology 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" (8, p. 1411). Many other courts have taken similar positions (10, p. 519; 18, p. 199; 32, pp. 1361–2; 33, p. 1385; 34, p. 969). Statements like this make one wonder how there was any understanding of cause and effect prior to the first large scale epidemiologic studies of the 1950s. As we have noted, for substances such as asbestos or thalidomide, causal associations were made in the absence of epidemiologic 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 relations, but not in every case.

Two-fold risk required

The two misuses of epidemiology discussed above involve courts excluding evidence that scientists consider relevant to evaluating causal relations by giving more weight to epidemiologic studies than do scientists. There is, however, another variation on the theme. Courts also exclude epidemiologic evidence that scientists value by giving some epidemiologic studies less weight than do scientists.

Thus, courts increasingly insist that epidemiologic studies must meet artificial, non-scientific standards of "scientific validity" in order to be heard by a jury. Perhaps the most dangerous of these court-created criteria is the requirement that any epidemiologic 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 of association" revealed in a population-based study with the probability that a substance has caused disease in a particular individual. By creating 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 (35). The derivation of individual "probabilities" of causation from population-based incidence data has been roundly criticized by epidemiologists (36). Yet it retains force in the judicial arena, and informs the requirement in certain cases that an epidemiologic study have a relative risk (or odds ratio) of 2.0 for a court to find admissible an expert's testimony based on such a study.

For instance, a federal district court judge in Maryland, in a case involving allegations that use of a contraceptive intrauterine device caused pelvic inflammatory disease, essentially equated the need to show a twofold increased risk with the requirement that the plaintiff prove causation by a preponderance of the evidence (37, p. 1092). In a case alleging injury from exposure to silicone breast implants, a district court in Oregon stated that plaintiffs must be able to show that exposure to breast implants more than doubled the risk of developing the alleged injuries (9, p. 1403). Another breast implant case heard in federal court in Texas offers a particularly telling example of management of scientific testimony by the courts. The court found that it was unreasonable as a matter of law (applying Daubert) for an expert to rely on a study with a relative risk of 1.49, a 95 percent confidence interval of 0.97, 2.28, and a p value equal to 0.067, to support an opinion that silicone breast implants caused a particular disease, where the study's authors identified confounding factors that weakened their confidence in the outcome of the study (38, p. 878). The court also stated that experts could not reasonably rely on a study of 445 women (210 with implants) that found a relative risk of 2.43 (95 percent confidence interval between 1.29 and 4.37) between exposure to silicone breast implants and sicca symptoms, because of the small size of the study (38, p. 880). The court suggested that to study a condition with high background
Incidence rates (like dry eyes, a sicca symptom), a larger sample size was required (38, p. 880). We might all agree that the studies described above do not on their own demonstrate a causal association between exposure to silicone breast implants and disease, but there is no scientific basis for completely ignoring the two studies, as the court insisted any “reasonable” expert would do.

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 epidemiologic 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” (39, p. 1321). Building on the Ninth Circuit’s statements, a federal district court judge held that plaintiffs had to submit epidemiologic studies showing a twofold increase in risk in order to proceed in the massive Hanford Nuclear Reservation Litigation (6, p. 13). These two rulings, considered together, demonstrate how far from science the courts have strayed. If a qualified physician is prepared to testify to causation on the basis of a properly-conducted clinical evaluation, it is indefensible to exclude that testimony (and preclude the plaintiff from recovering) because the only epidemiologic studies conducted showed, for example, a relative risk of 1.92. Unfortunately, however, district courts outside of the Ninth Circuit’s jurisdiction have relied on that court’s decision to impose their own similar requirements (32, p. 1356; 40, pp. 1225–7).

An important example of epidemiology’s misuse in the courtroom is a case from the Supreme Court of Texas, *Merrell Dow Pharmaceuticals v. Havner* (28). The court 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” (28, p. 717). The court’s explanation for this “rational basis” demonstrates its inadequate understanding of epidemiology. 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.

In the, the requirement that any epidemiologic study presented to a jury report a twofold 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 almost always takes place, the plaintiffs are also introducing other evidence to prove causation—such as pathology, animal experimentation, molecular modeling, or case studies—then requiring the epidemiologic 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” (41, p. 676).

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, as the *Havner* hypothetical perfectly illustrates, some studies may only measure an increased incidence in the subset of cases which would have not 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) (36, p. 1185). 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. 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 be biased against the plaintiff (36, p. 1193). For instance, the *Havner* court does 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 5 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 twofold magnitude of association between determinant and disease are ignoring the great difficulty of estimating individual risk from population-based data. The basic premise of probability of causation “holds only if the individual is truly representative of the reference population” (35, p. 807). 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, p. 80). Given unmeasured genetic and environmental factors, this assumption is likely false (42, p. 81). An age-specific probability of causation estimate would depend on the unknown mechanism by which exposure causes a particular outcome and the unknown degree of heterogeneity in the background disease rate. The court-created rule that any expert testimony offered in support of causation be supported by an epidemiologic study with a greater than twofold magnitude of association ignores this reality. By refusing to allow an expert to base her testimony on evidence in addition to epidemiology, courts demonstrate their misunderstanding of the limitations of population-based studies.

The reasoning employed by courts in arriving at a requirement that an epidemiologic study demonstrate a twofold magnitude of association also ignores an important complexity of disease process. Individuals vary in their response to a given disease determinant, depending on many factors. For instance, tamoxifen citrate has been hypothesized to increase the risk of contracting breast and uterine cancer in younger women, while it may act as a preventive agent in older women. Depending on how data are stratified, a study that reveals no statistical association between tamoxifen citrate and uterine or breast cancer 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 biologic relation over-
looks the fact that the strength of a factor's effect on a population depends on the relative prevalence of its causal complements (43, pp. 10–11). In other words, 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 strength of the effect of a given factor on the occurrence of a given disease may change because the prevalence of its causal complements in various mechanisms may also change" (43, p. 11). There are several examples of accepted causal relations 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 (43, p. 24). Courts, however, with the encouragement of some scientists, continue to incorrectly assume that a larger magnitude of association always indicates a greater likelihood of causal relation (as well as the converse of that proposition). This position is a drastic oversimplification.

**Peer review required**

Another prevalent legal misuse of epidemiology is the insistence by some courts that experts cannot testify if their opinions—or the epidemiologic analyses on which they are based—have not been subjected to the peer review process. Peer review and publication has been called the "most important means of ensuring that an expert's methodology is sound" (44, p. 897). The Havner court, while not requiring peer review publication, announced its skepticism of scientific evidence not subjected to the rigor of peer review (28, p. 727). The peer review requirement does most of its work where experts are testifying as to reanalysis of published data (4, 29). Courts have been resistant to allow experts to introduce their reanalysis in the courtroom without first having published it in a scientific journal. This wariness increases where experts seek to reanalyze data from a study in which the study's authors disagree with the expert's ultimate conclusions (44–46).

While on its face the peer review requirement may seem like a reasonable requirement, it is not necessarily 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 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 is unreliable. 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 (7, p. 593). 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. Recognizing that the "actual practice of medicine" does not require it, at least two federal appellate courts, the Third Circuit and the Eighth Circuit, have refused to require that studies be published and peer reviewed to be reliable (47, p. 155) and concluded that "victims of a new toxic tort should not be barred from having their day in court simply because the medical literature" is not complete (48, p. 1209). Epidemiologists do not refuse to consider other epidemiologists' views simply because they are not published in peer reviewed journals. Neither should the courts.

**Requiring a p value of less than 0.05**

Another way in which courts misuse epidemiology is by treating a p value of 0.05 as a magic number for evaluating the reliability of an epidemiologic study. Misusing the language of science, they say that, because this number must be met to ensure "statistical significance," studies that fail to meet it simply cannot be considered (6, pp. 49, 54; 28, pp. 724–8). One district court described a p value of less than 0.05 as "the most common value used to establish significance and to say that an observed association is probably real" (8, p. 1452).

As epidemiologists know, however, 0.05 is an arbitrary number, with more historical than inferential value (43). The dichotomization of study results into "significant" and "nonsignificant" 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 "steams 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" (43, p. 187). 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 relation 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." 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.

**Confidence intervals must exclude 1.0**

Finally, some courts are stating 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 (11, p. 33). Many courts that take this position cite the Fifth Circuit's decision in Brock as support (49, p. 992; 50, p. 461; 51, p. 783). 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" (28, p. 723). According to the court, if the confidence interval includes
favor 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 epidemiologic 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 relations to study, how to study them, and whether to publicize the results. Some courts have recognized this problem (37, p. 1094).

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 epidemiologic 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 epidemiologic studies, and publish those that are favorable to their defense.

The long-term effect of the misguided restrictions on evidence described in this presentation 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.

THE CAUSES OF LEGAL MISUSE OF EPIDEMIOLOGY

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 epidemiologic study to demonstrate a twofold 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 jurors. 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 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 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 presentation, we will provide two anecdotes that reflect the means by which some companies, and their supporters, further their goals.

In November 1990, the National Judicial College, which is the leading educational institution for state judges, put on a multi-day course entitled "Toxic Torts for Judges." It was attended by a large number of state court judges from around the country. The course was paid for by the Chemical Manufacturers Association and the Chicago law firm of Phelan, Pope & John—which represents defendants in toxic tort cases. Instead of ensuring that these sources of funding did not influence the course's content, the National Judicial College offered both funders the opportunity to participate in planning the course. The defense law firm accepted the offer and the resulting course, as noted in the participants' evaluations, was extremely one-sided. The only instructor on the central topic covered by the course was the president of a service that helped defend toxic tort suits and included among its clients, according to its promotional materials "nearly every major insurer and most major self-insured corporations." When Trial Lawyers for Public Justice questioned the National Judicial College about this, the National Judicial College offered to let Trial Lawyers for Public Justice have input into the course content in the future if it paid $25,000 to co-sponsor it. We understand the National Judicial College has since changed its procedures.

More recently, and far more successfully, America's largest corporations helped launch a massive campaign to convince policymakers 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 fantastic claims of the willingness of judges and juries to rely on the "far fringes of science" to assess liability (53). Huber described juries (and some ignorant 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 (54). 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 (54, pp. 1707–22).

SUGGESTIONS FOR ACTION

We propose several possible ways to improve the law's use of science in evaluating causal relations. 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 Science is even launching a pilot program to help federal judges identify potentially helpful court-appointed experts. 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 (friend of the court) briefs. Amicus briefs, filed either while a decision is under reconsideration or, more commonly, 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 should work with legal institutions to establish more formal feedback processes and enhance legal understanding of science. The Administrative Office of the US 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.

REFERENCES