SCIENTIFIC EVIDENCE IN THE COURTROOM: ADMISSIBILITY AND STATISTICAL SIGNIFICANCE AFTER *DAUBERT*

by

Robert P. Charrow
David E. Bernstein

Foreword
by
Honorable Patrick E. Higginbotham
U.S. Court of Appeals for the Fifth Circuit

Introduction
by
James S. Todd, M.D.
Executive Vice President
American Medical Association

Preface
by
Frederick L. Webber
President
Chemical Manufacturers Association

WASHINGTON LEGAL FOUNDATION
WASHINGTON, D.C.

This monograph is one of a series of original papers published by the Legal Studies Division of the Washington Legal Foundation. Through this and other publications, WLF seeks to provide the national legal community with legal studies on a variety of timely public policy issues. Additional copies of this monograph may be obtained by writing to the Washington Legal Foundation, Publications Department, 2009 Massachusetts Avenue, N.W., Washington, D.C. 20036.

Other recent studies in the WLF monograph series include:

A Punitive Damages Primer: Legal Principles and Constitutional Challenges by Arvin Maskin and Peter A. Antonucci, Weil, Gotshal & Manges. Foreword by Alfred C. DeCrane, Jr., Chairman and Chief Executive Officer, Texaco, Inc. Introduction by George S. Frazza, Vice President and General Counsel, Johnson & Johnson, 1994. Library of Congress No. 93-061849.

Regulatory Takings Under The Fifth Amendment: A Constitutional Primer by Michael M. Berger, Berger & Norton. Foreword by Honorable Randall R. Rader, U.S. Court of Appeals for the Federal Circuit. Introduction by Tad Taube, The Woodmont Companies, 1994. Library of Congress No. 93-061850.

Of Mice And Men . . . Surviving Endangered Species Act Litigation by David T. Hardy, Esq. Foreword by Congressman Charles W. Stenholm (D-Tx.). Introduction by Governor Kirk Fordice, State of Mississippi, 1993. Library of Congress Catalog No. 93-060800.

"Extraordinary How Potent Cheap Music Is": The Case For Reforming Lawyer Advertising by John D. Stuckemeyer, Esq. Foreword by Honorable Kenneth W. Starr, Kirkland & Ellis. Introduction by C.T. ("Kip") Howlett, Jr., Esq., Vice President - Government Affairs, Georgia-Pacific Corporation, 1993. Library of Congress Catalog No. 93-060553.

FDA Criminal Enforcement: How To Prevent And Defend Against Liability by Steven M. Kowal, Burditt & Radzius, Chartered. Foreword by C. Manly Molpus, President and Chief Executive Officer, Grocery Manufacturers of America, Inc., 1993. Library of Congress Card No. 92-085237.

Legal Strategies and Responses to Activist Environmentalist Lawsuits by Frank B. Cross, Associate Professor of Business Regulation, University of Texas. Foreword by Senator Larry E. Craig. Introduction by Richard L. Stroup, Political Economy Research Center, 1992. Library of Congress Catalog Card No. 92-080727.

Library of Congress Catalog Card No. 94-60916

SCIENTIFIC EVIDENCE IN THE COURTROOM: ADMISSIBILITY AND STATISTICAL SIGNIFICANCE AFTER *DAUBERT*

by

Robert P. Charrow
David E. Bernstein

Foreword
by
Honorable Patrick E. Higginbotham
U.S. Court of Appeals for the Fifth Circuit

Introduction
by
James S. Todd, M.D.
Executive Vice President
American Medical Association

Preface
by
Frederick L. Webber
President
Chemical Manufacturers Association

WASHINGTON LEGAL FOUNDATION
WASHINGTON, D.C.

FOREWORD

by
Honorable Patrick E. Higginbotham
U.S. Court of Appeals for the Fifth Circuit

In 1901, Learned Hand published an article in the *Harvard Law Review* sketching the history of methods for bringing expert knowledge to trials (Learned Hand, *Historical and Practical Considerations Regarding Expert Testimony*, 15 HARV. L. REV. 40, (1901)). Ninety-three years later, we continue to struggle as new scientific methods and old methods enhanced by computers are accommodated. The persistence of these tensions reflects the reality that the admissibility of expert testimony lies athwart a long restless legal-fault line — the respective roles of jury and judge. As we have moved from the model that every man knows the law to the professionalization of the law, we have said that the judge determines the facts. The rules of evidence rest upon this division of function, and none has been more elusive than the definition of the roles of judge and jury concerning expert testimony. This Washington Legal Foundation monograph engages the current debate over *Daubert*, a debate the Supreme Court has wisely or unwittingly left to flourish in the lower courts.

Close examinations of judicial decisions and publications such as this are then important. For now, a great deal rests upon disciplines other than law, including, for example, a working knowledge of the primary rules of statistics. Courts over the past twenty years have been required to engage use of statistical methods in employment practices and voting rights cases. These cases often consist almost in their entirety of mathematical models. Yet judges and lawyers as a group do a poor job with numbers. They, for example, persist in their ignorance of the relationship between statistical significance at the .05 level and legal causation. Other groups are attempting to furnish education. The Federal Judicial Center and the National Center for State Courts have added statistics to their curricula for judges. This is encouraging. This debate, however, is larger than mastery of technical protocol. Nearly 15 years ago, I explained in a large civil trial case tried to me as a trial judge:

> Excursions into the new and sometimes arcane corners of different disciplines is a familiar task

> of American lawyers and its generalist judges. But more is afoot here, and this court is uncomfortable with its implications. This concern has grown with the realization that the esoterics of econometrics and statistics which both parties have required this court to judge have a centripetal dynamic of their own. They push from the outside roles of tools for "judicial" decisions toward the core of decision making itself. Stated more concretely: the precision-like mesh of numbers tends to make fits of social problems when I intuitively doubt such fits. I remain wary of the siren call of the numerical display and hope that here the resistance was adequate; that the ultimate findings are the product of judgment, not calculation.

Vuyanich v. Republic Nat'l Bank, 505 F. Supp. 224, 394 (N.D. Tex. 1980).

My view has not changed.

Messrs. Charrow and Bernstein offer a thoughtful and useful monograph. It is a helpful contribution to this important debate. I applaud their efforts. There is no substitute for hard work and study of these "intractable" problems.

INTRODUCTION

by
James S. Todd, M.D.
Executive Vice President
American Medical Association

Scientific knowledge, especially in medical aspects, is constantly evolving. The infinite variation in patients, diseases, and reactions to medications and treatments makes determination of causal relationships difficult at best. As in other areas, medical science is constantly confirming or dispelling medical consequences of treatment decisions. The landscape is strewn with studies purporting to be scientific, but unable to stand the scrutiny of the scientific method. In our litigious society the courts are increasingly being asked to settle scientific issues. To date, the track record has not been very good.

"Junk science" has entered the legal and medical lexicon, meaning bizarre studies, theories, and opinion sometimes have been admissible in liability cases. The pernicious consequences of "junk science" are not difficult to discern as products and procedures are withdrawn as a consequence of litigation failing to apply scientific methods of determination. The consequent withdrawal of potentially beneficial products and medical services can have an entirely unwarranted adverse effect on public health.

This Washington Legal Foundation monograph reviewing the *Daubert* decision and its subsequent interpretations is well done and should be the standard by which future cases of medical science in dispute should be determined. Fundamentally, *Daubert* gives courts substantial discretion to evaluate expert testimony based on scientific methods, hopefully assuring more precise determinations in the medico-legal process.

Daubert per se gave little specific instruction other than any "qualified expert" providing "scientific, technical or other specialized knowledge . . . may testify at a trial." Left undefined were "qualified expert" and "scientific." Nonetheless, as shown

in this monograph, courts are increasingly functioning as "gatekeepers" of scientific validity. Yet courts continue to be confused at times by conflicting "expert testimony" and scientific data.

Participants in litigation and the public in general should welcome this shift toward scientific objectivity. The adversarial process is ill-suited to settling scientific disputes where fact should reign supreme. The goal should be objective evidence supported by methodologically valid statistics, leaving the court to decide the validity eschewing untested theories, opinion, and minority findings.

At a minimum, the following criteria should be considered in judging what is acceptable scientific evidence:

- professional qualifications of witnesses, including recognition of the qualifications considered by the scientific world as valid;
- conclusions should be based upon scientific evidence versus theory-based speculation;
- scientifically accepted methodologies should be used to obtain data for conclusions;
- publication in peer-reviewed literature, therefore subject to scrutiny and analysis by experts in the field should be routine;
- acceptance by the scientific community that a consensus has been reached;
- avoidance of conjecture, opinion, hearsay, and anecdotal reports; and
- recognition that one cannot transfer *in vitro* results or animal studies to intact human beings. Studies must be done on intact humans.

These criteria will serve the courts well in their quest for scientific accuracy. Statistical significance alone does not

provide cause and effect, a point brought out by the authors. It takes repeated studies, usually by multiple authors, with statistically significant findings before one can safely conclude that cause and effect is likely and/or proven.

The authors correctly conclude, "[i]f attorneys and judges can master the relationship among relative risk, confidence level, and legal causation, they will be able to ensure that the only plaintiffs who successfully rely on epidemiological evidence in toxic tort are those who actually meet the law's requirement that a plaintiff prove that the defendant more probably than not caused injury." Without this mastery, public health will be the loser as "junk science" continues to do a disservice in the all-too-frequent quests to assign responsibility where none exists. Scientists and their beneficiaries everywhere should applaud and promote the tenets of the *Daubert* case.

PREFACE

by
Frederick L. Webber
President
Chemical Manufacturers Association

Our society today is confronted with complex legal issues. Many of these issues involve a question of causation. For example, what is the cause of a birth defect or a particular disease, such as cancer? What health risks arise from taking certain drugs, eating certain foods, or working at certain jobs? Does exposure to a specific chemical or other hazardous substance adversely affect human health or the environment?

The Chemical Manufacturers Association strongly believes that sound scientific principles should be used to evaluate these and similar questions. In our view, sound science is the best guarantee of sound decision making. I can think of few matters more important to the chemical industry than the consistent application of good science by all levels of government.

The use of sound science is particularly important in the courtroom. Causation — whether an exposure to a chemical has in fact caused an injury — is a fundamental aspect of tort law. Whether exposure caused an injury should be established by reliable medical or scientific evidence, not unsupported opinions of so-called experts. As a simple matter of fairness, judges should exclude claims that rest solely on subjective opinions and unsubstantiated theories.

This does *not* mean that novel scientific findings that diverge from commonly held beliefs ought to be rejected out of hand. We fully support the development of innovative science and its use in the courtroom when it is developed in accord with the scientific method. The scientific process speaks to the *process* by which the scientist reached a conclusion, not to the conclusion itself. Consequently the court is not required to make

value judgments about the legitimacy of the results, or to reject novel scientific conclusions merely because they are novel.

Our nation has now taken a significant step toward the advancement of sound scientific testimony. The U.S. Supreme Court in *Daubert* articulated the standard of review for expert evidence. This decision should play a major role in keeping "junk" science out of the courtroom.

The Court repudiated the "let it all in" philosophy advocated by the plaintiffs' attorneys. Instead, the Court held that trial judges are required to act as "gatekeepers" with respect to the admissibility of scientific evidence. The trial judge must assess whether the reasoning or methodology underlying the testimony is scientifically valid and whether the reasoning or methodology apply to the facts in the case. Plaintiffs' contention that experts can testify merely on the basis of their credentials was rejected.

This monograph provides readers with valuable insights on the *Daubert* decision. It also offers an excellent review of lower court decisions that have transpired since *Daubert*. Finally, it examines statistical significance as a basis for proof of causation.

Particularly interesting is the authors' review of decisions that have been announced since the *Daubert* decision. This review indicates that judges are taking seriously their role as "gatekeepers." The authors conclude: "Most federal courts are interpreting *Daubert* as giving them wide authority to restrict the scope of admissible scientific evidence in toxic tort litigation, and are using that authority aggressively."

This monograph makes an important contribution to the continuing debate regarding the admissibility of scientific evidence. It should be read by all those who are interested in the role of science in the courtroom.

ABOUT THE AUTHORS

Robert P. Charrow is an attorney with the Washington, D.C. law firm of Crowell & Moring, where he specializes in legal issues associated with health care, scientific misconduct, scientific research, and biotechnology. He represents health care providers, pharmaceutical companies, universities, and other entities with respect to a broad range of issues including torts, regulatory matters arising under Medicare, Medicaid, the Food, Drug, and Cosmetic Act, and research issues arising under the Public Health Service Act, the Bayh-Dole Act, and the Technology Transfer Act of 1986.

Prior to joining Crowell & Moring, he served as the Principal Deputy General Counsel for the U.S. Department of Health and Human Services, and as such, supervised the Health Care Financing Administration, the Office of Inspector General, the Public Health Service, and the Food and Drug Administration. He actively participated in formulating Department policy with respect to Medicare reimbursement, anti-kickback regulations, drug and device approvals, tort liability, and technology transfer. He was Chairman of the Task Force on Scientific Misconduct, and served as a member of the Senior Staff AIDS Coordinating Council, Attorney General's Task Force on Tort Reform, and Secretary's Task Force on Medical Liability.

Before joining the Department he served as Deputy Chief Counsel to Reagan-Bush '84, the president's authorized campaign committee and prior to that, he was a law professor at the University of Cincinnati College of Law, where he specialized in torts, and law and science.

He has written extensively in the areas of health care law, and science and the law and has published approximately 80 journal articles. He is also the co-author of a legal text scheduled to be published next year by Little Brown & Company.

Mr. Charrow received his law degree from Stanford University in 1969 and did his undergraduate work in physics at Harvey Mudd College in Claremont, California.

David E. Bernstein is an associate at Crowell & Moring, in Washington, D.C. Mr. Bernstein graduated from Yale Law School in 1991, and served as a clerk for Judge David A. Nelson of the United States Court of Appeals for the Sixth Circuit. Mr. Bernstein is co-editor, with Kenneth R. Foster and Peter W. Huber, of *Phantom Risk: Scientific Inference and the Law* (MIT Press 1993), and is the author of a forthcoming manuscript in the *Cardozo Law Review* on scientific evidence after *Daubert*. He has also written about scientific evidence and other legal topics for such publications as the *Yale Law Journal*, *The Review of Litigation*, *Science*, and *The Wall Street Journal*.

The authors thank Bert Black, Esq. and Dr. Robert Smith (University of Michigan) for their comments and criticisms, all of which were appreciated but not always heeded. The first author takes full responsibility for any mathematical errors or misstatements.

TABLE OF CONTENTS

I. INTRODUCTION....1

II. THE *DAUBERT* DECISION....3

III. TOXIC TORT DECISIONS RELYING ON *DAUBERT* TO LIMIT SCIENTIFIC EVIDENCE....8

A. Circuit Court Opinions....8

1. *Porter v. Whitehall Laboratories, Inc*....8

2. *Elkins v. Richardson-Merrell, Inc*....10

3. *Conde v. Velsicol Chem. Corp.*....10

4. *DeLuca v. Merrell Dow Pharmaceuticals*....11

5. *Thomas v. American Cyanamid Co*....12

B. District Court Opinions....13

1. *Wade-Greaux v. Whitehall Laboratories, Inc.*....13

2. *In re Joint Eastern and Southern Asbestos Litigation*....16

3. *Chikovsky v. Ortho Pharmaceutical Corp*....17

4. *Haim v. Secretary of the Department of Health and Human Services*....19

C. Conclusion....22

IV. *DAUBERT* AND STATISTICAL

SIGNIFICANCE 22

A. Hypothesis Testing and Statistical Significance.....24

B. Statistical Significance and the Plaintiff's Burden of Proof.....28

V. CONCLUSION 33

ENDNOTES.....35

APPENDIX A 39

APPENDIX B 41

GRAPHS 44

SCIENTIFIC EVIDENCE IN THE COURTROOM: ADMISSIBILITY AND STATISTICAL SIGNIFICANCE AFTER *DAUBERT*

by

Robert P. Charrow
David E. Bernstein

I.

INTRODUCTION

In June 1993, the Supreme Court decided *Daubert v. Merrell Dow Pharmaceuticals*,[1] the most important case involving the admissibility of scientific evidence in seventy years. The issue in *Daubert* was whether the common law *Frye* rule,[2] which dates back to 1923, survived the enactment of the Federal Rules of Evidence. The most important of the Federal Rules concerning scientific evidence is Rule 702, which states that any qualified expert who possesses "scientific, technical, or other specialized knowledge [which] will assist the trier of fact to understand the evidence or to determine a fact in issue" may testify at a trial. The Supreme Court held that Rule 702 did indeed supersede the *Frye* rule.

It would be a mistake, though, to view *Daubert* as little more than a rudimentary discourse on one aspect of the Federal Rules of Evidence. Instead, the Court was asked to confront

fundamental issues concerning the relationship between science and the law, particularly the way scientific thought should be integrated into judicial decision-making. Although the Court did not resolve many of the important nuances before it, it nonetheless attempted to reconcile some of the basic differences in the way the two disparate disciplines approach and resolve comparable problems. Given the extent to which scientific theories and thought pervade the legal system,[3] such a reconciliation was long overdue. The importance of *Daubert* was not lost on the scientific community, which submitted many *amicus* briefs.

This monograph is divided into three sections. The first summarizes the Court's opinion in *Daubert*, noting the general guidelines the Court established for screening scientific evidence. We note that the Court did not address many narrow issues, including whether statistical significance ought to be a talisman for the admissibility of epidemiological evidence. The second section focuses on *Daubert's* progeny — an unusually rich and rapidly expanding body of lower court opinions, which, in little more than a year, has transformed what to many was an equivocal opinion into a major impediment to the admissibility of questionable science. The final section confronts and resolves the issue of statistical significance. It will be demonstrated that as a matter of both law and mathematics, statistical significance, while perhaps not a pivotal factor in determining admissibility, is critical in ascertaining whether a plaintiff's scientific evidence of causation is sufficient to satisfy a plaintiff's legal burden of proof.

II.

THE *DAUBERT* DECISION

Daubert involved claims unsupported by published scientific studies that the morning sickness drug Bendectin causes human limb reduction birth defects. The district court held that summary judgment must be granted to the defendant, the manufacturer of Bendectin, because the plaintiffs had failed to present statistically significant epidemiological evidence proving a causal relationship between Bendectin and the type of birth defects suffered by the plaintiffs. The Ninth Circuit Court of Appeals affirmed, holding that the plaintiffs' theory that Bendectin causes birth defects is not generally accepted in the scientific community, and was therefore inadmissible under the *Frye* rule. Under *Frye*, novel scientific evidence is admissible only when it has received "general acceptance" in the relevant scientific community.

On *certiorari*, the petitioners argued, among other things, that the *Frye* rule was superseded by the Federal Rules of Evidence, and further, that epidemiological evidence showing an increased risk ought to be admitted even though the results were not statistically significant at the .05 level and even though the study was never published in a peer-reviewed journal.[4]

The Supreme Court's *Daubert* decision was technically a victory for the plaintiffs. The Supreme Court found that *Frye* was indeed superseded by the promulgation of Rule 702 of the Federal Rules of Evidence, and therefore vacated the Ninth Circuit Court of Appeals opinion favoring the defendant.

The Court, however, rejected the "let-it-all-in" philosophy advocated by the plaintiffs' attorneys, and instead emphasized that the federal district courts must act as

"gatekeepers" and exclude evidence that is not scientifically reliable. The Court remanded the case along with new criteria for determining the admissibility of scientific evidence under Rule 702.

While *Daubert* doomed the *Frye* rule, many of the considerations that led to the adoption of that rule are still apparent in the *Daubert* opinion. As one court has noted, "[t]he decision in *Daubert* kills *Frye* but resurrects its ghost."[5] Like the *Frye* rule, the new test looks to the standards of the scientific community in determining the admissibility of scientific evidence. In particular, *Daubert* holds that proffered scientific evidence must constitute "scientific knowledge" to be admissible under Rule 702. This requirement, according to the Court, establishes a standard of evidentiary reliability. "Evidentiary reliability," the Court held, means "trustworthiness," and depends on "scientific validity."[6]

The Court added that Rule 702 requires that proposed expert scientific testimony must "assist the trier of fact to understand the evidence or to determine a fact in issue." Proposed testimony must therefore have some scientific relevance to the issue at hand. The fact that a study may be scientifically relevant for one purpose does not necessarily mean that it is scientifically relevant for other purposes. For example, high-dose animal studies are arguably relevant for risk research, and could therefore be admissible in the context of litigation over Environmental Protection Agency regulations. The same studies, however, have questionable relevance for actually predicting harm to humans from low-dose exposure, much less in establishing that the particular substance at issue was more probably than not the cause of human injury. Under *Daubert*, such studies are therefore not admissible to prove causation.

Having focused on the importance of reliability and relevance, the Court promulgated the following guidelines for screening scientific evidence under Rule 702:

(1) The court should determine whether the theory or technique in question can be (or has been) tested.

(2) Peer review is a relevant, though not dispositive, consideration.

(3) The known or potential rate of error of the technique should be determined, as should the existence and maintenance of standards controlling the technique's operation.

(4) Widespread acceptance can be an important factor, particularly with regard to a venerable theory or technique.[7]

The Court cautioned that these enumerated factors do not constitute a definitive checklist or test, and that other factors may be taken into consideration as well.[8]

The Court also noted that Rule 702 is not the only rule of evidence governing the admissibility of expert testimony. The Court, quoting Judge Jack Weinstein, a leading proponent of strict scrutiny of scientific testimony, stated that expert testimony should be subjected to particularly exacting scrutiny under Rule 403. That rule provides that a court should exclude evidence when its probative value is substantially outweighed by its prejudicial effect on the jury.

The Court also noted that evidence admissible under Rule 702 may still be excluded under Rule 703. This rule, which allows an expert to rely on hearsay only if the facts or data are "reasonably relied upon by experts in the particular field in forming opinions or inferences on the subject,"[9] can be a back-door to the *Frye* rule. Arguably, scientists presenting novel testimony may only "reasonably rely" on facts or data if those facts or data were gathered through a generally accepted methodology.

Even if a party manages to persuade a court that its evidence is admissible, the case will not necessarily go to the jury. *Daubert* squarely holds that courts have both the right and duty to direct a verdict or grant summary judgment if a party's scientific evidence is admissible but "insufficient to allow a reasonable juror to conclude that the position more likely than not is true."[10]

Despite the Supreme Court's clear holding that scientific evidence must be both reliable and relevant to be admissible under Rule 702, and the Court's affirmance of the viability of independent barriers to admissibility under Rules 703 and 403, some commentators have insisted that *Daubert* establishes an extremely liberal test for the admissibility of scientific evidence. Prominent Bendectin plaintiffs' attorney Barry Nace, for example, has declared complete victory.[11] Professor Michael Green of the University of Iowa Law School agrees that "*Daubert* is a resounding defeat for the [anti-] `junk science' school."[12]

After some initial confusion, plaintiffs' attorneys, who stand to benefit most from a liberalization of evidentiary standards, have settled on an interpretation of *Daubert*. Their official line is that the decision allows a court to examine the principles and methodology that underlie the opinions of causation only to determine whether it is the "type of reasoning based on the type of principles that people in this field use to come to those kind of conclusions."[13]

This interpretation of *Daubert* is clearly incorrect. The fact that an expert has engaged in an accepted "type of reasoning" does not mean that his testimony constitutes scientific knowledge and will be helpful to the trier of fact as required by Rule 702 as interpreted in *Daubert*. A court, for example, must ensure that expert analysis does not contain methodological, mathematical, or logical flaws.

While the plaintiffs' bar clearly has an axe to grind, even some opponents of "junk science" have been skeptical about the

ultimate effect of *Daubert* on the admissibility of scientific evidence. They worry that the Supreme Court's eradication of the strict, simple, "general acceptance" test would encourage a natural inclination among judges to spare themselves the time and energy required to properly scrutinize scientific evidence, and lead them to adopt a "let-it-all-in" approach.

The debate over the meaning of *Daubert* has flourished because the Supreme Court failed to apply the standards it established to the particular facts before it in *Daubert*. Despite the district court's ruling that epidemiological studies that are not statistically significant are inadmissible, for example, the Court declined to address this issue at all. The Court's failure to give more concrete guidance, combined with the necessarily broad nature of its dicta, understandably encouraged wide variations in interpretation.

However, there are now a sufficient number of post-*Daubert* opinions to conclude that the plaintiffs' attorneys and the skeptics have been proven wrong. Most federal courts are interpreting *Daubert* as giving them wide authority to restrict the scope of admissible scientific evidence in toxic tort litigation, and are using that authority aggressively. As of this writing, nine federal courts, including five circuit court panels, have relied upon *Daubert* in either excluding proffered scientific evidence as to causation entirely, or finding that the evidence was insufficient as a matter of law.[14] In contrast, only two courts have relied upon *Daubert* in finding expert scientific testimony as to causation of injury to be admissible or sufficient.[15] The next section of this monograph will describe the nine decisions in the toxic tort context which have relied on *Daubert* to find expert testimony excludable or insufficient.

III.

TOXIC TORT DECISIONS RELYING ON *DAUBERT* TO

LIMIT SCIENTIFIC EVIDENCE

A. Circuit Court Opinions

1. *Porter v. Whitehall Laboratories, Inc.*, 9 F.3d 607 (7th Cir. 1993).

In this case, the Seventh Circuit Court of Appeals was faced with the claim that the decedent's ingestion of ibuprofen caused a kidney condition which advanced to rapidly progressive glomerulonephritis (RPGN), a kidney disease which led to decedent's death. The district court had excluded the plaintiff's scientific evidence and granted summary judgment to the defendant.

Two of the plaintiff's experts were the decedent's treating physicians. Dr. Diane Wells testified that ibuprofen did lead to the decedent's demise, but added, "What I'm giving you now is a kind of curb side opinion. If . . . you were asking me to give you an analytical, scientific opinion, then, I would have to research it, and I have neither the time nor the inclination to do that." The Seventh Circuit upheld the exclusion of her testimony, noting that "a `scientific' opinion is just what Rule 702, as interpreted in *Daubert*, requires."

Another of the plaintiff's physicians, Dr. Richard Combs, also testified. He admitted that he could not state to a reasonable degree of scientific certainty that ibuprofen caused the plaintiff's death, nor could he point to studies, records, or data on which he based his opinion. The court therefore upheld the exclusion of his testimony, because it was not well-grounded in the scientific method as required by *Daubert*.

The court also rejected the testimony of Dr. Fred Ferris. Dr. Ferris theorized that ibuprofen aggravated an independently

developed kidney problem, ultimately causing the decedent's demise. However, Dr. Ferris admitted that such an aggravation would be dose-related and would require a far greater dosage than the decedent ingested.

A fourth doctor, Dr. Francesco Del Greco, admitted that he based his testimony on animal tests which led to a "hypothesis, the proof of which remains to be made." The court rejected this testimony as speculative and not scientifically relevant to the issue at hand.

The plaintiff also presented testimony from Dr. David Benjamin, a pharmacologist, who wished to testify that the decedent's ingestion of ibuprofen had caused RPGN. Dr. Benjamin admitted, however, that in order to analyze the cause of the decedent's kidney failure, it would be necessary to rule out other possible causes. He also admitted that he did not know what those other causes were, so he could not rule them out. The court held that whatever the validity of Dr. Benjamin's theory that ibuprofen can cause RPGN, his inability to discount alternative causes in the particular case before the court meant that there was no "fit" between the theory and the case before the court, as required under *Daubert*.

2. *Elkins v. Richardson-Merrell, Inc.*, 8 F.3d 1068 (6th Cir. 1993).

This case, like *Daubert*, involved an allegation that the morning sickness drug Bendectin caused the plaintiff's birth defects. Before *Daubert*, the Sixth Circuit had twice upheld summary judgment for the defense in cases involving Bendectin. In *Elkins*, the court noted that in *Daubert* the Supreme Court cited one of those two cases[16] with approval. The *Elkins* court added that *Daubert* indicated that even if expert opinion or evidence on one side is admissible, summary judgment for the other side is appropriate if the evidence is not sufficient to create a genuine issue of material fact. The court held that *Elkins* was factually indistinguishable from the Sixth Circuit's two earlier Bendectin cases, and therefore affirmed the district court's grant of summary judgment to the defense.

3. *Conde v. Velsicol Chem. Corp.*, 1994 LEXIS 10752 (6th Cir.)

In *Conde*, plaintiffs alleged that Velsicol's commercial termiticide was the cause of their family's health problems and the loss of their property's value. The termiticide was negligently applied to the Conde's home by a third-party exterminator.

The court held that expert opinion must be both admissible and sufficient to justify a denial of a motion for summary judgment. *Conde*, like *Elkins*, held that the insufficiency of the plaintiff's expert medical testimony was ample justification to uphold summary judgment for the defendants. The case is distinct from *Daubert* in that it examines this separate requirement, rather than determining only the circumstances under which such testimony is admissible.

The court, relying on *Turpin v. Merrell Dow Pharmaceuticals, Inc.*, 959 F.2d 1349 (6th Cir. 1992) held that

even admissible testimony, if it "would not allow a jury to conclude by a preponderance of the evidence that the Condes suffered personal injuries," may be the basis for a grant of summary judgment. This does not contravene the holding in *Daubert*, since the Supreme Court in that case "construe[d] *Turpin* to treat the plaintiff's expert opinion indicating a basis of support for the plaintiff's theories . . . to be admissible but simply inadequate . . . ." *Conde*, 1994 LEXIS at 5 (quoting *Elkins*, 8 F.3d at 1071). *Conde* assumes the plaintiff's expert testimony to be admissible, but bases its decision on the sufficiency of the evidence to withstand summary judgment for Velsicol.

The court combined factors used in previous holdings to determine that the expert testimony was insufficient to permit a jury to conclude that exposure to the pesticide caused the harm suffered by the family. Plaintiff's experts were unable to eliminate other potential causes; their theories were incompatible with medical tests performed on the Condes' tissue samples; the ideas proposed were inconsistent with most peer-reviewed scientific literature; and the tests used were experimental and not generally accepted by the scientific community. Thus, the Sixth Circuit held that "general acceptance" remains an important criterion not only for the admissibility of scientific evidence, but for its sufficiency as well.

4. *DeLuca v. Merrell Dow Pharmaceuticals*, No. 92-5287 (3d Cir. Aug. 4, 1993), *aff'g*, 791 F. Supp. 1042 (D.N.J. 1992).

DeLuca is yet another case involving allegations that Bendectin caused a plaintiff's birth defects. In an initial district court opinion, the court excluded the plaintiff's expert testimony and granted summary judgment. On appeal, the Third Circuit, in an opinion that helped cement its status as the most liberal federal court of appeal on the issue of admitting scientific evidence, held that the district court had not sufficiently justified its exclusion of

plaintiff's experts' testimony. The court therefore vacated the district court's decision, and remanded the case for further consideration.

On remand, the district court engaged in an exhaustive analysis of plaintiff's proffered expert evidence, and once again ordered it excluded and granted summary judgment for the defense. The plaintiffs once again appealed to the Third Circuit.

Before the Court of Appeals could render its decision, *Daubert* was released. Both parties submitted supplemental briefs to the court, arguing that *Daubert* supported their respective positions on admissibility.

Daubert apparently led some judges on the Third Circuit to change their stance on the issue. Rather than reversing again, the court affirmed the district court's grant of summary judgment in a cursory, one-paragraph opinion that did no more than cite *Daubert*. *DeLuca* is perhaps the best available evidence that *Daubert* will be interpreted as a "strict scrutiny" opinion by the vast majority of lower federal courts.

5. *Thomas v. American Cyanamid Co.*, 1993 LEXIS 24470 (6th Cir.).

In *Thomas*, the plaintiff alleged that the defendant's DPT vaccine caused the plaintiff's temporary injuries and aggravated his pre-existing brain abnormality. The district court granted summary judgment to the defendant on the first issue, and a directed verdict on the second. On appeal, the Sixth Circuit, in a summary opinion, upheld the district court's finding that the plaintiff's evidence was insufficient as a matter of law. The court noted that the "trial court's ruling on the sufficiency of the plaintiff's expert evidence is not inconsistent with *Daubert*, which focuses on the admissibility of expert evidence."

B. District Court Opinions

1. *Wade-Greaux v. Whitehall Laboratories, Inc.*, 1994 WL 80840 (D.V.I.).

Plaintiff Jacqueline Wade-Greaux brought this product liability action on behalf of her daughter. The plaintiff alleged that her use during pregnancy of the over-the-counter asthma medication Primatene Mist and Primatene Tablets caused her daughter to be born with birth defects.

Following the demands of *Daubert*, the court conducted a hearing spanning seven separate days regarding the admissibility of the plaintiff's experts' testimony. The court required the plaintiff's witnesses to address both general causation and specific causation. The court explained that general causation concerns whether the agent at issue is capable of causing birth defects in humans at therapeutic dose levels, while specific causation concerns whether that agent caused the particular malformations found in the particular plaintiff.

The court proceeded to assess the reliability of the plaintiff's experts' testimony under criteria set forth in *Daubert* and two Third Circuit cases, *United States v. Downing*, 753 F.2d 1224 (3d Cir. 1985), and *DeLuca v. Merrell Dow Pharmaceuticals, Inc.*, 791 F. Supp. 1042 (D.N.J. 1992), *aff'd*, 6 F.3d 778 (3d Cir. 1993). Synthesizing these three cases, the court listed five factors to be considered when evaluating the reliability and soundness of a particular methodology. It then proceeded to evaluate the experts' testimony with regard to these factors:

a. The novelty of the methodology and its relationship to more established methodologies accepted by the scientific community.

The court concluded that the relevant scientific community is teratologists, and that their accepted methodology is that exposure to an agent during pregnancy should be associated with an increased frequency of a distinctive pattern of birth defects, as shown through repeated, consistent human epidemiological studies. This element was absent from the respective methodologies of each of the plaintiff's experts.

b. The existence of specialized literature.

The court found no evidence that any of the methodologies advanced by the plaintiff's experts had been subjected to peer review, or that any specialized literature existed endorsing their methodologies.

c. The non-judicial uses to which the scientific technique is put.

The court found that none of the plaintiff's experts had presented to the scientific community the theory that each of them offered for litigation purposes — that Primatene Tablets and Mist can cause birth defects in humans at therapeutic dosage. One of the plaintiff's experts, Dr. Gilbert, had expressed the opinion to her scientific peers that Primatene may be teratogenic in humans. The court held, however, that an opinion regarding a mere possibility of general causation does not meet the relevant criterion.

d. The qualifications and professional stature of the expert witnesses employing the methodology.

The court found that four of the plaintiffs' five witnesses do not, as part of their regular activities, study the causes of birth

defects in humans. The fifth expert, meanwhile, relied upon a flawed, unpublishable study.

e. The frequency with which a technique leads to erroneous results.

The court first discussed the experts' reliance on *in vivo* and *in vitro* animal studies. The court found that "[t]he notion that one can accurately extrapolate from animal data to humans to prove causation without supportive positive epidemiological studies is scientifically invalid." The court noted that the plaintiff's experts each acknowledged that different species react differently to the same agent, that at some dosage virtually any substance is teratogenic in an animal species, and, finally, that different routes of administration affect the teratogenic impact of an agent.

The plaintiff's experts had also relied on individual human case reports, Drug Experience Reports, and other anecdotal evidence. The court concluded that "such data represent anecdotal information of chance associations, do not purport to assess cause and effect and have no epidemiological significance."

Having reviewed all of the plaintiff's experts' evidence, the court concluded that their opinions about the teratogenic potential of the products at issue in humans at therapeutic doses amounted to "rank speculation and conjecture." The court therefore excluded this testimony, and granted summary judgment to the defense.

2. *In re Joint Eastern and Southern District Asbestos Litigation*, 827 F. Supp. 1014 (S.D.N.Y. 1993).

In this case, the surviving spouse of a sheet metal worker brought an action against asbestos products manufacturers and

contractors asserting that the worker's fatal colon cancer was caused by his exposure to asbestos. The jury found in favor of the plaintiff, and the defendant filed a post-verdict motion seeking to overturn that decision.

The plaintiff had presented expert testimony regarding statistically significant epidemiological studies purporting to show that there was a sufficiently strong relationship between asbestos exposure and colon cancer that the plaintiff's cancer could be attributed to such exposure. The court stated that all epidemiological evidence must be examined in the context of Hill's criteria. The court then proceeded to carefully assess whether the plaintiff's evidence met the criteria.

First, the court reviewed the strength and consistency of association, concluding that the various epidemiological studies relied on in the plaintiff's proof of causation "establishes only the conclusions that the association between exposure to asbestos and developing colon cancer is, at best, weak, and that the consistency of this purported association across the studies is, at best, poor." Next, the court examined the dose-response relationship between asbestos and colorectal cancer, concluding that it was "erratic, at best."

The court then analyzed experimental evidence — animal studies of potential pathological changes in animals after exposure to asbestos. The court found that the experimental evidence "fail[ed] to establish any causal relationship between the exposure to asbestos and the development of cancer in animals." The court followed with a discussion of the plausibility criterion, finding that the relationship between exposure to asbestos and colorectal cancer is nothing "more than possible."[17] A mere possibility, the court held, does not satisfy the plausibility criterion of sufficiency.

Finally, the court discussed whether the plaintiff's epidemiological evidence met the coherence criterion. The court noted that colon cancer has various known confounding

conditions, and that asbestos is not considered to be a risk factor for colon cancer, while various other factors are recognized in the medical literature such as a high-fat and/or low-fiber diet, hereditary syndromes, and other confounding factors. The plaintiff presented clinical evidence that he was at no special risk from these factors. The court dismissed this evidence as a "superficial differential diagnosis" and found that the coherence criterion also was not met.

Having found that none of Hill's criteria that the court examined were met, the court reversed the jury's verdict and granted judgment for the defendant on the ground that the evidence was insufficient as a matter of law under *Daubert*.

3. *Chikovsky v. Ortho Pharmaceutical Corp.*, 832 F. Supp. 341 (S.D. Fla. 1993).

While pregnant with Honey Chikovsky, Sara Chikovsky applied Retin-A twice daily to her face and neck for the treatment of acne. Honey Chikovsky was subsequently born with a variety of birth defects. Honey's parents sued, alleging that Sara's use of Retin-A during pregnancy caused Honey's birth anomalies. The plaintiffs relied solely on the opinion of their expert witness, Dr. Bertram, in support of this allegation. Dr. Bertram testified that in his opinion Retin-A is a teratogen, and that it caused Honey's birth defects. The defendant then moved for summary judgment.

After a review of *Daubert*, Judge Kenneth Ryskamp proceeded to rule that Dr. Bertram's testimony was inadmissible for the following reasons:

a. Dr. Bertram did not rely on any published material in forming his opinion that the topical application of Retin-A causes birth defects. In fact, he admitted that he was not aware of any published article or treatise which reports any

study that has found that Retin-A causes birth defects.

b. Dr. Bertram's theory had not been tested. There was a total lack of data linking Retin-A to birth defects.

c. Dr. Bertram testified that the dose of a particular substance is relevant in determining whether it acts as a teratogen. He had no knowledge, however, of how much Retin-A Sara Chikovsky absorbed through her skin while pregnant.

d. Dr. Bertram testified that he based his opinions concerning Retin-A on studies regarding the teratogenic effects of high doses of vitamin A and other vitamin A derivatives. Dr. Bertram testified, however, that he prescribes prenatal vitamins, which contain Vitamin A, to his pregnant patients. He also testified that he did not know at what dosage level Vitamin A became unsafe for use by pregnant women. Most significant, according to the court, was that Dr. Bertram had not compared the doses of vitamin A in the studies he relied upon to the doses found in Retin-A.

e. Dr. Bertram relied on studies showing that Accutane, another acne medication derived from Vitamin A, is teratogenic. But he admitted that there are not enough studies on Retin-A to determine whether the birth defects associated with Accutane are also associated with Retin-A. The analogy drawn between the two drugs is therefore wanting.

f. Dr. Bertram failed to consider whether there were genetic explanations for Honey's birth defects.

g. Finally, Dr. Bertram testified that he relied on his "common sense" in determining that Retin-A is a teratogen. The court, however, noted that under *Daubert* scientific knowledge connotes more than a subjective belief or unsupported speculation. "This is precisely the kind of evidence that the trial judge must exclude in performing the gatekeeper function."

The court concluded that Dr. Bertram's opinions were not based on scientifically valid principles, and therefore did not meet the reliability requirements of Rule 702. Dr. Bertram's opinion on causation was therefore deemed inadmissible, and the court granted summary judgment to the defense.

4. *Haim v. Secretary of the Department of Health and Human Services*, 1993 LEXIS 145 (Cl. Ct.).

In *Haim*, the plaintiff alleged that a DPT vaccination caused the death of her daughter, Nicole, from neurological problems. Nicole had suffered a seizure five days after being vaccinated and suffered encephalopathy (injury to the brain). Because the plaintiff brought the case under the National Childhood Vaccine Act, the Federal Rules of Evidence did not apply. Nevertheless, the court found that *Daubert's* discussion of what criteria should be used in determining the credibility of scientific evidence to be instructive, and relied upon those criteria in its opinion.

The plaintiff relied upon the testimony of Drs. Mark Geier and Gerald Slater. These experts relied primarily on an epidemiologic study proposing that the DPT vaccine could cause seizures up to seven days after it was administered. Dr. Geier also relied upon certain animal tests. The court found that this

testimony was not persuasive, and granted summary judgment to the government. The opinion is somewhat confused, and no model of judicial reasoning. Wading through the opinion, however, it seems that the court got the science right, and granted summary judgment for the right reasons.

The court found that the epidemiologic study relied upon by the doctors was fundamentally flawed. The basis for its conclusions rested upon merely seven cases. These children were presumed to be normal at the time of vaccination, but no prevaccination neurological testing had been performed. Two of the seven children were not diagnosed with encephalopathy, but with seizures. Of the other five children, three had other conditions that may have caused their problems. That leaves two cases of encephalopathy and two cases of seizures out of 1,182 cases of serious acute neurologic illness in children ages two to thirty-five months as the basis for the conclusion that DPT can cause neurologic damage up to seven days after vaccination, a dubious extrapolation.

Moreover, the authors of the study later noted that their study had not been replicated by other case-control studies and does not demonstrate that the DPT vaccine causes permanent brain damage. They also admitted to various flaws in their methodology, such as defining the date of onset in the study as the first onset of acute neurological symptoms rather than the onset of any symptoms. This may have lengthened the possible interval between vaccination and neurological harm.

In sum, the court concluded, the study at best showed a possible association between the DPT vaccine and neurological injury in a small number of children. But, found the court, "it is inconceivable that any scientist could justifiably reach a conclusion" on individual causation from the study.

The court also rejected Dr. Geier's reliance on animal studies. According to Dr. Geier, pertussis (which is in the DPT vaccine) causes brain damage partly because it causes endotoxins

to enter the lungs. He also claimed that use of endotoxins on animals has produced illness in the test subjects. The court, however, found that because Dr. Geier did not describe these studies with any specificity, the court could not accept them as the basis for a valid opinion on causation in fact.

Moreover, added the court, there are methodological deficiencies in animal testing which mar its acceptability as proof of causation. In animal tests, endotoxin itself, not DPT vaccine, is injected into laboratory animals for testing purposes. Because Dr. Geier did not describe the difference between the amount of endotoxin injected into the animals and the highest amount that could possibly be found in a particular lot of DPT vaccine, the animal tests could not be used to prove causation. Dr. Geier also failed to account for the difference in effect between an endotoxin injected numerous times (or in places unsuitable for humans, such as directly into the brain) into laboratory animals for the purpose of making them sick, and a DPT vaccine containing some amount of endotoxin injected once into a child. The court concluded that there "are too many variables here . . . to conclude DPT's effect on humans is analogous to endotoxin's effect on animals."[18]

The court concluded with an attack on Dr. Geier's credibility. According to the court, Dr. Geier is a `hired gun' who "has made a career of testifying in cases involving long-onset encephalopathy following DPT vaccine." The court concluded that in the wake of *Daubert*, "no other court should be without the tools with which to dissect Dr. Geier's testimony and to recognize its frailty."

C. Conclusion

These post-*Daubert* cases lead to an impression that *Daubert* will have a tremendous positive impact on scientific evidence in toxic tort cases. Courts have used *Daubert* to engage in generally sophisticated and comprehensive reviews of the pertinent scientific evidence. Most important, they have recognized that *Daubert's* demand that scientific evidence be both reliable and relevant requires them to crack down on a wide range of speculative, unreliable "junk science" evidence, including evidence based *post hoc ergo propter hoc* reasoning, clinical testimony that does not take dosage into account, unreliable epidemiological studies and animal studies. Perhaps most important, even when courts find that shaky evidence is admissible, *Daubert's* emphasis on the courts' gatekeeper function has encouraged them to engage in thorough reviews to ensure that evidence is sufficient to support causation.

IV.

DAUBERT AND STATISTICAL SIGNIFICANCE

For defense attorneys facing unreliable scientific evidence of causation, *Daubert* is proving to be a useful tool, but they face the heavy burden of properly explaining the relevant scientific issues to the courts. Despite their generally correct interpretations of *Daubert*, the opinions discussed above are confused at times. For example, the court in *In re Asbestos Litigation* simply asserted that the plaintiff's faulty epidemiological and other evidence was admissible, without explaining why.

The Court of Federal Claims' opinion in *Haim*, meanwhile, indulges in a rather confusing monologue

differentiating between scientific and legal standards of causation. The Special Master in that case made the common error of confusing the commonly-used scientific standard of statistical significance (.05) with a scientific standard of cause. The same confusion was especially apparent during oral argument in *Daubert* when Professor Gottesman, counsel for the Petitioner, argued that it made little sense to exclude evidence that was not statistically significant at the 0.05 level since the preponderance of the evidence standard only requires certainty at the 0.5 level.

The fact that a study purports to show an association between a substance that the plaintiff was exposed to and the injury he suffered, and is statistically significant at the .05 level, does not prove that there was a 95% chance that the substance caused the type of injury injury suffered by the plaintiff. Assuming that the study was methodologically sound, it merely means that there was only a one in twenty chance that the purported association was an "outcome of chance."[19]

However, few (if any) epidemiological studies are perfect, and many have grave flaws. Before an epidemiological study is accepted as evidence of legal causation, it must first be analyzed for obvious methodological flaws. To ensure that subtle biases did not creep in, the study also must meet some of Hill's criteria before it can be considered even remotely reliable.

Even a study that does show an association does not necessarily show a *causal* association. For example, there is an association between silicone breast implants and *reduced* rates of breast cancer. This does not mean that breast implants caused the reduction in breast cancer.[20] Rather, it probably means that women whose natural breast size is relatively small (and are therefore more likely to get breast implants) have a lower risk for breast cancer than women whose natural breast size is relatively large (and are therefore less likely to get breast implants).[21] Thus, breast implants are associated with, but might not cause, reduced rates of breast cancer.

Thus, the petitioner's argument that statistical significance should not be the be all and end all of admissibility is correct, but for the wrong reasons. Whether the results of a proffered study are statistically significant is one factor among others that a court may properly consider under *Daubert* in determining the admissibility of scientific evidence.

Statistical significance also has an extremely important role to play in determining whether a plaintiff's evidence of causation is sufficient to withstand a summary judgment motion.

Below we discuss statistical significance, its role in hypothesis testing, and its relationship to the plaintiff's burden of proof.

A. Hypothesis Testing and Statistical Significance

The Court in *Daubert* emphasized that one of the factors that a trial court should consider in determining admissibility is whether the proffered conclusions are capable of "falsification." This was a short hand way of asking whether the proffered conclusions are subject to experimental testing, in the scientific sense. Simply stated, scientists can rarely "prove" through experimentation that their hypotheses are correct. Instead, experiments are performed to see if hypotheses can be disproved directly (by finding a counter-example) or proved indirectly by setting up an alternative hypothesis that the researcher then seeks to disprove. As scientists perform more and more experiments, and all of them fail to disprove the hypothesis, the scientific community may eventually accept the validity of that hypothesis.

This inability to directly prove a hypothesis is especially apparent in the biological and social sciences. Studies of human biology strive to draw conclusions by observing effects on a representative sample of the general population. Studies of this

type involve inherent uncertainty because they attempt to draw conclusions relevant to the general population based on an examination of a subset of that population.

Suppose, for example, that a researcher decides to test a hypothesis that a given drug is associated with a specific well-defined side effect in humans. It is clearly impossible to test this hypothesis using all humans; a sample will have to suffice. Accordingly, the researcher recruits 200 human subjects and randomly divides them into two groups of 100 each. Subjects in the first group receive sugar pills (placebo) while those in the second group receive the drug.[22] Neither the subjects nor the experimenter knows who is receiving the drug and who is receiving the placebo. In terms of hypothesis testing, the researcher's null (H_o) hypothesis (the one that he or she hopes to disprove or nullify) is

$$H_o: \quad T_{se}=C_{se},$$

where T_{se} is the probability that a person suffers the side effect given that that person is in the treatment group and C_{se} is the probability that a person suffers the side effect given that that person is in the control group.

The set of hypotheses that remains viable if we reject the null hypothesis is called the alternative hypotheses. One alternative hypothesis, for instance, could be that $T_{se}>C_{se}$.

If we reject the null hypothesis and it turns out that we were wrong, then we have committed what is called a type one error. In the context of our experiment, a type one error occurs if we conclude that the drug causes more side effects than the placebo, when in fact it does not. On the other hand, we commit a type two error if we accept the null hypothesis (the drug does not cause more side effects), when in fact it really does. The level of statistical significance is nothing more than the upper bound of the probability of committing a type one error, *i.e.*, rejecting the null hypothesis when it is in fact true. *See* Winer, Statistical

Principles in Experimental Design 11 (2d ed. 1971). As the level of statistical significance decreases (*e.g.*, from 0.05 to 0.01), the probability of committing a type two error increases.

Let's now return to our experiment. Suppose that, after the subjects have taken their regimen, the researcher examines them and notes that 10 subjects in the treatment group (drug group) have developed the side effect while only 5 in the control group (placebo) have developed the side effect. In other words, the relative risk ratio is 2.0 (*i.e.*, 10/5), that is to say, subjects in the treatment group are twice as likely to have developed the side effect than subjects in the control group. However, are we willing, based on this one experiment, to conclude that the drug increases the risk of the side effect?

It could be that pure chance and not the drug was responsible for the observed increase in the side effects. Specifically, in drawing conclusions from the data, we are assuming that the 100 subjects in the control group represent the general population and that therefore the likelihood that the side effect will spontaneously occur in the general population is only 5%. It is possible, though, that the control group was not a representative sample and that the actual frequency of the side effect in the general population is 10%. If that were the case, then there would be no increase in the relative risk associated with the drug. It is also possible that the treatment group is not representative and that while the actual rate of a spontaneous side effect in the general population is only 5%, it occurs with a greater frequency (10%) in the treatment group.

Statistical testing allows determination of the likelihood that our observations are due to pure chance and not to differences in the way that the two groups are treated. When a scientist reports that his or her results are statistically significant, what it means is that the likelihood that the observed differences were due to chance is less than some predetermined probability, which by custom has been set at 0.05. In other words, results are statistically significant if there is less than a 5% chance that the

observed differences are due to pure chance. As it turns out, in our hypothetical study above, there is an 18% chance of observing a relative risk ratio of 2.0 given that the real relative risk ratio is 1.0. Since 0.18 is much greater than 0.05, our observed differences are not statistically significant. For example, if the experiment were performed 100 times, but instead of giving half the subjects the drug, all the subjects were given sugar pills, we would expect to observe the same differences in about 18 runs of the experiment as observed in the actual experiment using the drug.

There has been much debate among respected scientists concerning the importance of statistical significance and whether it is proper to reject the null hypothesis where the probability of a type one error is greater than 5%. One factor that may lead some scientists to use a higher level of statistical significance (*e.g.*, 0.1) is that they may be concerned about committing a type two error, *i.e.*, accepting the null hypothesis when it is false.

Statistical significance is certainly an important factor that scientists take into consideration in drawing conclusions. However, it is an arbitrary level set by custom. It does come into play under *Daubert*, indirectly, through, for example, the "publication in a peer reviewed journal" factor. Peer reviewed journals may be unwilling to publish conclusions where the level of statistical significance departs markedly from the 0.05 level. Also, studies with a sufficiently relaxed level of statistical significance (*e.g.*, 0.1) may be so speculative as to not be sufficiently reliable to be admissible under *Daubert*.

B. Statistical Significance and the Plaintiff's Burden of Proof

In mass tort cases, the mere fact that a plaintiff is able to present a statistically significant study showing a causal association between the substance at issue and the type of injury suffered by the plaintiff does not necessarily mean that the plaintiff is able to present a prima facie case of causation. Rather, the plaintiff must prove legal causation, namely that it is more probable than not that his or her exposure to the substance caused the adverse health effect. In epidemiological terms, a plaintiff must show a causal association sufficiently high as to prove that his exposure to the substance more than doubled his risk of being injured. As a practical matter, this means that a plaintiff must first establish "general" causation, namely that there is an established association between the agent and the type of injury or disease suffered by the plaintiff. Second, the plaintiff must then bridge the gap between general causation and legal causation by introducing some evidence that would enable the finder of fact to conclude that results of epidemiological studies can be applied to the plaintiff. If the plaintiff is unable to show "general causation" or cannot show specific causation, then judgment should be entered for the defendant.

To illustrate, the experiment described above is modified so that there are now two groups, each consisting of 1000 subjects. As before, those in the control group receive sugar pills while those in the treatment group receive the drug believed to be associated with some adverse health effect. The researcher observes that 80 subjects in the treatment group develop the side effect, as compared to 50 subjects in the control group. The relative risk ratio is thus 1.6 (*i.e.*, 80/50). Owing to the larger sample size, this difference is in fact statistically significant at the 0.01 level (*i.e.*, there is less than 1% likelihood that the observed difference is due to pure chance).

Even though the results are statistically significant, they may be insufficient to withstand a motion for summary judgment. As noted above, a plaintiff must prove by a preponderance of the evidence that the drug or other agent caused injury. Here, all the plaintiff can show is that 80 subjects who had been given the drug developed the ailment, but that 50 subjects in the control group also developed the ailment. In other words, of the 80 subjects, 50 of them would have developed the ailment even if they had not received the drug. That means that there is only a 30 in 80 chance (*i.e.*, 37.5% chance) that the drug "caused" the ailment or injury. This would be insufficient as a matter of law to satisfy the preponderance of the evidence standard (greater than 50% likelihood that the drug caused the adverse effect exhibited by the plaintiff), even if the plaintiff were one of the subjects in the study which is almost never the case. Therefore, unless a plaintiff has additional evidence (*e.g.*, evidence to show that he or she is unusually susceptible), the defendant would be entitled to a summary judgment.

It is for this reason that several courts have held that an epidemiological study can only establish a prima facie case of causation if it shows a relative risk of greater than 2.0.[23] However, from a mathematical perspective, the relative risk needed to satisfy the preponderance of the evidence standard increases as the level of statistical significance increases (*e.g.*, from 0.05 to 0.10). In other words, while a RR of 2.2 may be sufficient to get to the jury if the results are statistically significant at the 0.05 level, a higher RR would be needed if the level of significance were 0.15. In short, while the level of statistical significance may be little more than artifact or custom within the scientific community and therefore, subject to change by those drawing scientific conclusions, it is, for mathematical reasons, an integral part of the judicial decision-making which cannot be ignored.

To provide an *intuitive* view for the relationship between statistical significance and the minimum relative risk ratio needed to satisfy the preponderance of the evidence standard, assume an

epidemiological study is undertaken which reveals that the relative risk associated with a chemical agent is 2.1. There is, however, inherent uncertainty associated with our relative risk ratio. Given that we are dealing with samples, we actually do not know what the real relative risk ratio is. It could be larger or smaller than 2.1. Specifically, if we performed our experiment an infinite number of times and then plotted the resulting risk ratios, they may fall along some form of bell-shaped curve. Such a curve is called a "probability density function." There is a large family of bell-shaped curves that could describe our data, including the familiar normal curve (*i.e.*, gaussian probability density function) and the less common gamma density function. The width of the curve provides a measure of variability and is related to statistical significance. The wider the curve, the greater the standard deviation. To illustrate this point, we have plotted at Graphs 1 and 2, two possible normal curves, each with average relative risk ratios of 2.1. The normal curve in Graph 1 has a standard deviation of 0.7, while the normal curve in Graph 2 has a standard deviation of 0.4. For an <RR>=2.1 to be statistically significant at the 0.05 level, it must be separated from an <RR>=1.0 (*i.e.*, no increase in relative risk) by about two standard deviations. Since the standard deviation in Graph 1 is 0.7, the average risk ratio of 2.1 would not be statistically significant since it is separated from <RR>=1 by about 1.58 standard deviations. In contrast, the average relative risk ratio of 2.1 in Graph 2 would be statistically significant since it is separated from an <RR>=1 by well over two standard deviations (*i.e.*, (2.1 - 1.0)/0.4).

As noted above, one minus the inverse of the "real" relative risk yields the probability that the plaintiff would not have experienced the side effect but for the drug or other chemical agent. For example, if the real relative risk is 4.0, the probability of legal causation is 0.75 (*i.e.*, 1 - (1/4)). If the real relative risk is 2.0, the probability of legal causation is 0.5 (*i.e.*, 1 - (1/2)). However, due to the inherent uncertainty one can never know the real relative risk.[24] Instead, a researcher may, after having conducted a number of experiments, have some idea about the

average relative risk and even some idea as to the nature of the probability density function associated with that relative risk. Indeed, for any given experiment, as the number of subjects increases, the observed relative risk will approach the mean relative risk. But the mean relative risk ratio is only one parameter. There is a variance associated with that mean, as illustrated in Graphs 1 and 2. When one takes into account that variance (or uncertainty), we find that the probability of general causation is always less than one minus the inverse of the average relative risk ration (*i.e.*, 1 - (1/<RR>)). In other words, one minus the inverse of the average relative risk ratio forms the upper bound of the probability of general causation. A mathematical proof of this proposition is attached at Appendix A-1. In other words, if someone reports having observed a <RR> of 2.0, the probability of legal causation based on the experiment is *always less* than 0.5. Correspondingly, a RR which is slightly greater than 2.0 will rarely if ever be sufficient to satisfy the preponderance of the evidence standard.

The notion that the probability of general causation always lags behind 1 - (1/<RR>), has profound implications with respect to statistical significance. As the level of statistical significance increases, the probability of causation decreases. This can be illustrated by using a bell-shaped curve known as the gamma probability density function. If our relative risk obeys a gamma function, then

$$P(C) = 1 - \frac{<RR>}{<RR>^2 - \sigma^2}$$

where P(C) is the probability of general causation, <RR> is the average or mean relative risk ratio and σ is the standard deviation (See Appendix B). As can be seen, as the standard deviation goes to zero, then P(C)=1-(1/<RR>). Conversely, as the standard deviation increases, then P(C) decreases even though the average relative risk ratio remains the same! This means that if we have two curves with identical average risk ratios, but one is statistically significant at the 0.05 level and the other is

statistically significant at the 0.1 level, the curve that is statistically significant at 0.05 will have a smaller standard deviation than the other curve. Correspondingly, the probability of causation will be larger for the curve that is statistically at the 0.05 level than it would be for the curve that is statistically significant at the less rigorous 0.1 level.

When the average relative risk ratio is said to be 2.1, for instance, and is statistically significant at the .05 level, what is really being said is that there is less than a 1/20 chance that the mean real risk ratio is less than or equal to 1.[25]

If experiments yielded an average relative risk ratio of 2.75 (for a gamma density function) with a statistical significance of 0.05, the probability of general legal causation is only about 52%, barely enough to withstand a motion for summary judgment. However, if the data were statistically significant at the 0.10 instead of the 0.05 level, then the probability of legal causation drops to about 43%, which is insufficient to withstand a defense motion for summary judgment. The relationship between statistical significance and the probability of legal causation for a gamma function is illustrated in Graph 3.

These trends are valid for any probability density function. However, they may not be as dramatic if our data obeyed a normal bell-shaped curve. Nonetheless, these examples and the proofs attached at Appendix A and Appendix B demonstrate that the level of statistical significance, while arbitrary for purposes of drawing scientific conclusions, plays a pivotal and deterministic role in ascertaining whether a plaintiff has met his or her burden of proof with respect to general causation. Ironically, therefore, while scientists may safely view the issue of statistical significance as of philosophical interest only, the courts do not have that luxury. Owing to the mathematics and the plaintiff's burden of proof, statistical significance cannot be ignored by courts or attorneys.

V.

CONCLUSION

As we have seen, *Daubert* established strict standards for the admissibility of scientific evidence. The standards, however, are also quite broad and somewhat vague. In general, courts are properly applying *Daubert* to exclude evidence based on animal studies, anecdotal evidence, and other types of dubious evidence. Other courts are holding that even potentially admissible evidence simply is not always sufficient to overcome a summary judgment motion by the defense.

Perhaps the most vexing issue the courts have yet to confront in a meaningful way is the relationship between statistical significance and *Daubert*. Rather than focusing on whether an epidemiological study must be statistically significant at the .05 level to be admissible, or whether a slightly higher level can be allowed, courts should focus on the weight that should be accorded to epidemiological evidence proffered to prove causation. As demonstrated in this monograph, a plaintiff must generally establish an average relative risk of more than 2 before his epidemiological evidence can establish a prima facie case of causation. As our level of statistical significance increases (*e.g.*, from .05 to .1), the higher the required relative risk.

If attorneys and judges can master the relationship among relative risk, confidence level, and legal causation, they will be able to ensure that the only plaintiffs who successfully rely on epidemiological evidence in toxic tort litigation are those who actually meet the law's requirement that a plaintiff prove that the defendant more probably than not caused his injury.

ENDNOTES

1. 113 S. Ct. 2786 (1993).

2. *Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923).

3. *See, e.g.*, Kenneth R. Foster, PHANTOM RISK: SCIENTIFIC INFERENCE AND THE LAW (1993).

4. The actual study that the petitioner tried to have admitted was not really a "study," but rather a reanalysis of data collected by other researchers. These other researchers found no relationship between Bendectin and limb reduction birth defects.

5. *In re Joint Eastern & Southern District Asbestos Litigation*, 827 F. Supp. 1014, 1033 (S.D.N.Y. 1993).

6. *Daubert*, at 113 S. Ct. at 2795 n.9. Two articles relied upon by the Court in addressing the issue of scientific validity provide further explication of this issue. Bert Black, *A Unified Theory of Scientific Evidence*, 56 FORDHAM L. REV. 595 (1988); Starrs, *Frye v. United States Restructured and Revitalized: A Proposal to Amend Federal Evidence Rule 702*, 26 JURIMETRICS J. 249 (1986).

7. *Daubert*, 113 S. Ct. at 2796-7.

8. *Id.*

9. *Id.* at 2797.

10. *Id.* at 2798; *Accord In re Joint Eastern & Southern District Asbestos Litigation*, 827 F. Supp. 1014, 1050 (S.D.N.Y. 1993).

11. Barry Nace, *Reaction to Daubert*, Shepard's Expert and Scientific Evidence Q., July 1993, at 51.

12. Michael D. Green, *Relief at the Frying of Frye: Reflection on Daubert v. Merrell Dow Pharmaceuticals*, Shepard's EXPERT AND SCIENTIFIC EVIDENCE Q., July 1993, at 43, 45.

13. *E.g.*, Ron Simon, *High Court Throws Out Rigid Rules Excluding Scientific Evidence, Says Focus Must Be on Methods, Principles*, BNA Product Safety & Liability Reporter, Summer/Fall 1993, Special Report on *Daubert*, at 10.

14. *Elkins v. Richardson-Merrell, Inc.* 8 F.3d 1068 (6th Cir. 1993); *Thomas v. American Cyanamid Co.*, 1993 LEXIS 24470 (6th Cir.); *Porter v. Whitehall Laboratories, Inc.*, 9 F.3d 607, (7th Cir. 1993); *DeLuca v. Merrell Dow Pharmaceuticals*, No. 92-5287 (3d Cir. Aug. 4, 1993), *aff'g*, 791 F. Supp. 1042 (D.N.J. 1992); *Wade-Greaux v. Whitehall Laboratories, Inc.*, 1994 WL 80840 (D.V.I.); *Chikovsky v. Ortho Pharmaceutical Corp.*, 832 F. Supp. 341 (S.D. Fla. 1993); *In re Joint Eastern and Southern District Asbestos Litigation*, 827 F. Supp. 1014 (S.D.N.Y. 1993); *Haim v. Secretary of the Department of Health and Human Services*, 1993 LEXIS 145 (Cl. Ct.).

15. *Cantrell v. GAF Corp.*, 999 F.2d 1007 (6th Cir. 1993); *Leary v. Secretary of Department of Health and Human Services*, 1994 WL 43395 (Fed. Cl.).

16. *Turpin v. Merrell Dow Pharmaceuticals, Inc.*, 959 F.2d 1349 (6th Cir. 1992).

17. *Id.* at 1046.

18. The court also criticized Dr. Geier for ignoring an Institute of Medicine Study categorically rejecting any

interpolation of animal studies in the context of attributing causation of neurologic illness to DPT vaccine.

19. For ease in reading the text, we are using the phrase "outcome by chance," instead of the more technically correct definition of statistical significance. More accurately, statistical significance is defined in terms of the probability of obtaining a result that leads us to reject a hypothesis, when in fact that hypothesis is true. If an observed difference is statistically significant at the .05 level what we are really saying is that there is a less than one in twenty chance that the difference might still have been observed even if there was really no difference at all. For example, suppose we are presented with an urn that contains millions of balls, some of which are red and some black. We are not permitted to look inside the urn. Nevertheless, we believe and thus, hypothesize that half the balls are red and half are black. We then are permitted to draw ten balls from the urn and find that all ten are red. The likelihood of drawing ten red balls, if in fact half were red and half were black, is less than 1/1000. Given this outcome, we should reject our hypothesis.

20. H. Berkel, *Breast Augmentation: A Risk Factor for Breast Cancer*, 326 NEW ENG. J. MED. 1649 (1992).

21. *Id.*

22. This hypothetical study would probably be unethical to perform. *See* 45 C.F.R. Part 46.

23. *DeLuca v. Merrell Dow Pharmaceuticals, Inc.*, 911 F.2d 941, 958-59 (3d Cir. 1990); *In re Joint E. & S. Dists. Asbestos Litig.*, 827 F. Supp. at 1027; *Marder v. G.D. Searle & Co.*, 630 F. Supp. 1087, 1092 (D. Md. 1986), *aff'd*, 814 F.2d 655 (4th Cir. 1987); *Cook v. United States*, 545 F. Supp. 306, 308 (N.D. Cal. 1982).

24. In fact most researchers when reporting relative risk use a confidence interval (*e.g.,* we are 95% certain that the real relative risk is between 0.8 and 5.0).

25. Given the nature of the alternate hypothesis we are using a one-tailed test for statistical significance. In other words, an observed risk ratio is deemed statistically significant if there is a greater than 95% chance that the real risk ratio is above 1.0.

APPENDIX A

Appendix A-1

Assume that the relative risk ratio can be described by a random variable X, such that

$$P(-C) = P(-C \mid x_1)p(x_1) + P(-C \mid x_2)p(x_2) + \ldots$$

$$P(-C) = (1/x_1)p(x_1) + (1/x_2)p(x_2) + \ldots$$

$P(-C) = \Sigma(1/x_j)p(x_j)$, where $p(x_j)$ is the probability mass function and $p(x=0)=0$. Assume that $p(x)$ is a probability density function[1] defined over the domain $[0,\infty]$. Then,

$$P(-C) = \int(1/x)p(x)dx.$$

In Appendix A-2, it is proven that for any probability density function

$$\int(1/x)p(x)dx > 1/\int(x)p(x)dx$$

Therefore, $P(C) = 1 - \int(1/x)p(x)dx < 1 - 1/(\int(x)p(x)dx)$

Thus, $P(C) < 1 - 1/\langle RR \rangle$, for any probability density function.

---

[1] A probability density function has certain properties, including that $\int p(x)d(x) = 1$ over the domain. For a more complete description of these properties *see* E. Parzen, *Modern Probability Theory and Its Applications* 151 (1960).

Appendix A-2

The proof below, developed by a graduate student at Carnegie Mellon University, is based on one that appeared in W. Rudin, *Real and Complex Analysis*, 3rd ed., McGraw-Hill Book Co., New York (1987), where Rudin proves a form of Jensen's inequality relating to convex functions.

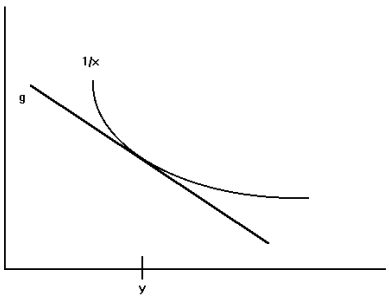
Define g to be to the curve (1/x) in the below graph at point y. Because 1/x is convex, g lies below 1/x. Note that g(y) = 1/y and the slope of 1/x to the right of y, and greater than the slope of 1/x to the left of y. Let m be the slope of 1/x at y (that is, $m = -1/y^2$). Then g is given by g(x) = m(x-y) + 1/y.

So, if y = _xp(x) dx, we get _(1/x)p(x)dx > _g(x)p(x)dx =_(m(x-y) + 1/y)p(x)dx

= m_xp(x)dx - my_p(x)dx + 1/y _p(x)dx

= my - my + 1/y = 1/(_xp(x)dx) QED

It should be noted that there are other more sophisticated proofs for the above proposition, but those have been omitted in the interests of brevity.



APPENDIX B

To prove that for the gamma density function, $P(C) = 1 - \langle RR \rangle / \langle RR \rangle^2 - \sigma^2$

where $\langle RR \rangle$ is the average or expected relative risk, σ is the standard deviation,
and $\sigma < \langle RR \rangle$.

$P(C) = 1 - \int_0^{\infty} 1/x\,p(x)dx$, where $p(x)$ is the gamma density function.

$$\int_0^{\infty} p(x) = 1.$$

$$P(C) = 1 - \int (1/x)(\lambda/\Gamma(r))(\lambda x)^{r-1} e^{-\lambda x} dx$$

$$= 1 - \lambda^r (1/\Gamma(r)) \int x^{r-2} e^{-\lambda x} dx$$

Integrating by parts:

$du = x^{r-2}$ $\quad v = e^{-\lambda x}$

$u = x^{r-1}/r\text{-}1$ $\quad dv = -\lambda e^{-\lambda x}$

Thus,

$$P(C) = 1 - \lambda^r (1/\Gamma(r)) \{ x^{r-1}/r\text{-}1\, e^{-\lambda x} \Big|_0^{\infty} + \int \lambda/(r-1) x^{r-1} e^{-\lambda x} dx \}$$

$$= 1 - \int \lambda/\Gamma(r)\, (\lambda x)^{r-1} (\lambda/r\text{-}1) e^{-\lambda x} dx$$

$= 1 - (\lambda/r\text{-}1) \int_{\lambda}/\Gamma(r) (\lambda x)^{r-1} e^{-\lambda x} dx$

$= 1 - (\lambda/r\text{-}1)$

But for the gamma function we know that $\langle x \rangle = r/\lambda$ and $\sigma = (r)^{.5}/\lambda$, therefore it follows that

$P(C) = 1 - \langle x \rangle / \langle x \rangle^2 - \sigma^2$

QED