PESSIMISM ABOUT MILWARD

MICHAEL D. GREEN†

Courts in the toxic tort arena have been struggling for two decades with the task, imposed by Daubert v. Merrell Dow Pharmaceuticals, Inc.,1 of determining which expert causation opinions meet the reliability standard imposed by Daubert.2 Milward has been hailed as a fresh way to approach the rigors of determining causation in toxic tort litigation.3 Unlike many prior cases, it recognizes that scientists use a weight of the evidence methodology to make causal inferences and that no algorithm exists to guide this determination.4 It also recognizes that reasonable experts can disagree and that rather than an objective scientific assessment, scientists must employ judgment and interpretation in making causal determinations.5

This article addresses the genesis and development of the Comment in the Restatement (Third) of Torts that influenced the Milward opinion. Looking forward, this paper expresses pessimism about the potential that Milward holds for reforming the way in which courts employ Daubert to determine the admissibility of an expert opinion on causation. The essence of the problem is the different contexts in which experts operate in adversary-driven litigation and in their professional roles.

† Williams Professor of Law, Wake Forest School of Law.
2. Id. at 592–93 (requiring “a preliminary assessment of whether the reasoning or methodology underlying the testimony is scientifically valid and of whether that reasoning or methodology properly can be applied to the facts in issue”).
5. Id. at 22.
I. INTRODUCTION

I must confess that I was as delighted as anyone else when I first read Milward v. Acuity Specialty Products Group, Inc. Milward was the first court to explicitly accept and endorse the scientific methodology of weight of the evidence in deciding a Daubert motion. It did so, in the face of numerous courts that took a Balkanized approach to the scientific evidence, examining each study separately and determining whether it was adequate to support the challenged expert’s opinion.

The reasons for my favorable reaction to Milward go back to a several-year period leading up to the American Law Institute’s (“ALI”) annual meeting in May 2005. For those who are unfamiliar with the Restatement (Third) of Torts, it has a lengthy Comment that addresses proof of causation in toxic substances cases, such as was the issue in Milward. Comment c, as it has come to be called, was the most difficult and controversial provision in this piece of the Restatement (Third) of Torts, as

---

6. Others participating in this Symposium have done a fine job of explaining Milward, so I will not subject the reader to another recitation of the case.
7. Joseph Sanders identified several other cases that have employed the weight of the evidence methodology. Joseph Sanders, Milward v. Acuity Specialty Products Group: Constructing and Deconstructing Science and Law in Judicial Opinions, 3 WAKE FOREST J.L. & POL’Y 141, 158–59 (2013). The two cases that take weight of the evidence most seriously in the toxic causation context fall well short of the ringing endorsement found in Milward. In King v. Burlington Northern Santa Fe Railway Co., 762 N.W.2d 24 (Neb. 2009), the weight of the evidence discussion by the court occurred in the course of a lengthy and general exposition about epidemiology and the scientific method. The issue in King was whether plaintiff’s expert could testify to causation where there were conflicting epidemiologic studies on the question. Id. at 32–33. Weight of the evidence played almost no role in resolving that question. In In re Seroquel Products Liability Litigation, 2009 WL 3806435, at *6 (M.D. Fla. 2009), the legitimacy of weight of the evidence as a method was not in dispute; the parties clashed over whether plaintiff’s expert witness properly applied it in forming her opinion on causation.
9. The Restatement (Third) of Torts is being prepared in several subject-specific volumes. Products Liability (1998) and Apportionment of Liability (2000) have already been published. The project to which I refer is titled Liability for Physical and Emotional Harm, the first volume of which was published in 2010 (the second volume was published in 2012). The provisions contained in there to which I refer in this paper were finally approved at the ALI annual meeting in 2005. 82nd Annual Meeting, 2006 A.L.I. PROC. 92 (2006) [hereinafter PROCEEDINGS 2005].
explained below. Although Comment c received final approval in 2005, other provisions approved with it in 2005 had influenced and been adopted by courts. Despite its notoriety during the restating process, Comment c languished and received almost no judicial attention until Milward. But Milward repeatedly cited, quoted, and appeared to rely on Comment c in adopting a weight of the evidence methodology, i.e., the idea that scientists so engaged were required to use judgment and could reasonably reach different conclusions based on the same body of scientific evidence. As any scholar will attest, there is no greater reward than having your research make an impact.

There are two principal matters I address in this article. The first is the controversy over Comment c that nearly prevented it from ever seeing the light of Restatement day. The second is to explain why, despite my elation in reading Milward with its virtually unique and sophisticated understanding of scientific methodology, after further reflection, I am pessimistic that it will be a harbinger for better decision making about toxic causation and the admissibility of testimony from experts who opine on that subject.

II. COMMENT C

During the Restatement drafting process, some took the view that there should not be any treatment of the subject of causation in toxic substances litigation. Initially, it was the plaintiffs’ interests that opposed Comment c, because it too much reflected the impact of the Daubert revolution, and plaintiffs were still licking the wounds inflicted on them by that dramatic reform. 


12. See, e.g., Thompson v. Kaczinski, 774 N.W.2d 829, 839 (Iowa 2009); A.W. v. Lancaster Cty. Sch. Dist. 0001, 784 N.W.2d 907, 918 (Neb. 2010).


14. Sanders, supra note 11, at 1030–32.

15. And we should be clear that the reform entailed far more than moving from a Frye general acceptance standard to the Daubert four-factor informed reliability standard. As I have written elsewhere, Frye played almost no role in screening the substance of expert’s opinions in civil litigation until the emergence of toxic substances litigation in
Daubert challenges to expert testimony often left plaintiffs without evidence for the factual causation element of their prima facie case, thus resulting in dismissal when a plaintiff’s expert’s testimony was declared inadmissible.\(^\text{16}\)

Plaintiffs also objected to Comment c because it developed a three-element test for factual causation that, in more conventional tort cases, seemed to require but a single element.\(^\text{17}\) For parties with the burden of proof, more elements mean more ways in which to run afoul of those proof burdens and thereby suffer dismissal on summary judgment or as a matter of law.\(^\text{18}\) Unsurprisingly, defense lawyers defended Comment c and opposed modifications to it.\(^\text{19}\)

Plaintiff dissatisfaction with Comment c was evident at the 2002 annual meeting of the ALI. Dr. Anthony Robbins of the Tufts University School of Medicine attended that meeting as a guest.\(^\text{20}\) During the discussion of Comment c, Dr. Robbins was granted the floor.\(^\text{21}\) The doctor criticized the draft of Comment c, stating that several statements were “clear, in my opinion, misstatements of...science” or reflected a misunderstanding of scientific principles that “leaves everyone in doubt as to whether you know what you are talking about...”\(^\text{22}\)

Considering Dr. Robbins’s remarks, I wondered which is worse: misunderstanding scientific principles or misstating science? But his remarks had a profound impact because, unlike controversies over the ethical obligations of lawyers to clients, the thin-skull rule, or the rule against perpetuities, members of the

---

\(^{16}\) See, e.g., Claar v. Burlington N. R.R. Co., 29 F.3d 499 (9th Cir. 1994); Daubert v. Merrell Dow Pharm., Inc., 43 F.3d 1311 (9th Cir. 1995).

\(^{17}\) Comment c contains subsections addressing exposure, general causation, and specific causation. Each is necessary to establish the existence of agent-disease causation. Joseph Sanders explains why this conceptualization is necessary in most toxic substances cases and not different from more traditional traumatic-injury causation requirements. See Sanders, supra note 11, at 1030–32.


\(^{21}\) Id.

\(^{22}\) Id. at 294.
ALI had no expertise on which they could draw to evaluate Dr. Robbins’s claim that requiring a relative risk of two is scientifically incorrect, or his claim that increasing the number of subjects in an epidemiology study can identify small effects with “an almost indisputable causal role.”

Nevertheless, we felt that it was important to press on and have the Third Restatement address the issue of proof of causation in disease cases. Professor Joseph Sanders, one of the country’s leading academic experts on the subject, observed that Comment c addresses the most important development in the law of causation over the past half-century. For a Restatement of Torts to take a bye would be most unfortunate, especially given that scientific principles, with which lawyers and judges are not familiar, are enmeshed in legal doctrine.

23. As one member of the ALI put it: “I think people in the audience could be excused for thinking, when we start talking about § 28, Comment c, that they have wandered into the wrong door of The Mayflower [the hotel where the meeting was held] and suddenly were in a scientific meeting and not a law meeting.” Id. at 316 (statement of Patrick A. Malone).

24. I responded to Dr. Robbins in a letter dated June 12, 2002, explaining that the requirement of a relative risk of two to satisfy specific causation is not a scientific concept but a legal one developed because epidemiology does not address the matter of the cause of an individual’s disease. See Michael D. Green et al., Reference Guide on Epidemiology, in REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 549, 608–09 (3d ed. 2011) (“[E]pidemiologic studies do not address the question of the cause of an individual’s disease. This question, often referred to as specific causation, is beyond the domain of the science of epidemiology [and] . . . is a necessary legal element in a toxic substance case.”). Criticizing the relative risk threshold as a misunderstanding of scientific principles is a little like criticizing ice cream because it can’t be used to clean carpets.

Robbins’s second claim was equally incorrect. Epidemiology, unlike clinical trials, is observational rather than experimental. When case and control groups cannot be randomized and exposure depends on subjects’ voluntarily having exposed themselves to a suspected toxin, two important sources of error, in addition to random error, are injected: bias and confounding. See LEON GORDIS, EPIDEMIOLOGY 183–89 (discussing bias and confounding); Philip Cole, Causality in Epidemiology, Health Policy, and Law, 27 ENVTL. L. REP. 10279, 10281 (1997). Increasing the number of study participants, as Robbins advocated, only reduces the risk of random error and has no impact on bias or confounding. Robbins joins the Fifth Circuit Court of Appeals, which made the same mistake in Broek v. Merrell Dow Pharmaceuticals, Inc., 884 F.2d 166 (5th Cir. 1989), asserting that statistical methods for assessing random error could negate bias and confounding. See Michael D. Green, Expert Witnesses and Sufficiency of Evidence in Toxic Substances Litigation: The Legacy of Agent Orange and Bendectin Litigation, 86 NW. U. L. REV. 643, 667–68 (1992).

25. The plural usage is not royal, but includes my co-reporter, Bill Powers. It is safe to say that I, who had spent considerable time researching and writing about toxic causation before taking on Restatement work, was more persuaded of the need to persevere on Comment c than he was.

At that time, a member of the ALI, Patrick Malone, suggested that the Institute work with the National Academy of Sciences (“NAS”) to vet the draft of Comment c.\(^\text{27}\) Leaders of the ALI immediately recognized the benefits of that course, and, with the cooperation of the Science, Technology, and Law Program of the NAS, a panel of five eminent epidemiologists and physicians was assembled to review a draft of Comment c.\(^\text{28}\) A meeting took place at the NAS in Washington, D.C. with representatives of the ALI.\(^\text{29}\)

We did not learn anything revolutionary, but there were some points made by the scientists that were valuable. The scientists told us that they do not think in probabilistic terms about causation.\(^\text{30}\) So the legal standard of proof, a preponderance of the evidence (or more likely than not), is not one that scientists use in their professional lives unless they have the misfortune of testifying as an expert.\(^\text{31}\) Instead, the standard of proof used in scientific decision-making depends on the comparative costs of false positives and false negatives.\(^\text{32}\) Thus, if nontreatment of a serious disease will result in the patient’s death, a treatment that has an exceedingly low likelihood of success may be employed.\(^\text{33}\) They did tell us about the use of the weight of the evidence methodology when assessing evidence that bears on whether a

\[\text{References}\]

29. A transcript of the meeting can be found at http://sites.nationalacademies.org/PGA/stl/PGA_049555 (last visited Nov. 6, 2012).
31. Their fate as expert witnesses today is actually improved somewhat with the reform of the threshold standard for an expert to testify “to a reasonable degree of medical or scientific certainty.” Prompted by a drumbeat of critical commentators, courts have begun to recognize that physicians and scientists do not use or have any idea what a reasonable degree of certainty is and have reformulated the threshold for expert testimony to the nearly universal civil justice standard of a preponderance. See Jeff L. Lewin, The Genesis and Evolution of Legal Uncertainty About “Reasonable Medical Certainty,” 57 Md. L. Rev. 380, 384 (1998).
33. See Peter A. Glassman et al., The Role of Medical Necessity and Cost-Effectiveness in Making Medical Decisions, 126 Annals Internal Med. 152, 152–54 (1997).
causal relationship exists. That weight of the evidence methodology entails gathering all available evidence relevant to the causal issue and making a judgment in light of that evidence. They also explained that using the Sir Bradford Hill criteria to assess whether an association found in a study was truly causal or spurious required informed judgment—no algorithm exists for making that determination. Thus, reasonable experts can differ in their judgments about such matters.

We took the points that the scientists provided us and built them into Comment c. Much of what we did looked better to plaintiffs’ lawyers and having the draft vetted by the NAS did wonders for the confidence of the ALI in the work we were producing on a topic that only a handful of members understood.

But still there was resistance. The evidence scholars accused us of turf grabbing—after all, the battles over proof of causation had turned into challenges to the admissibility of expert testimony under Daubert. The evidence folks thought this was now adjectival law, and they didn’t want the torts Barbarians trespassing on their land. What those evidence scholars failed to appreciate is that Daubert determinations in toxic substances cases have, to a large extent, evolved to an assessment of the sufficiency of the evidence proffered by plaintiffs’ experts to support an opinion on causation. That is how Judge Thomas Penfield Jackson handled the matter in an early Bendectin case, Richardson v. Richardson-Merrell, Inc., until the federal courts took a wrong turn, in my judgment, and began a misguided expert admissibility

34. Goodman, supra note 32, at 34.
37. See Douglas L. Weed, Underdetermination and Incommensurability in Contemporary Epidemiology, 7 KENNEDY INST. ETHICS J. 107, 107 (1997).
approach. That Daubert hid the sufficiency-of-the-evidence ball with an inapt, multifactorial reliability approach should not obstruct the Restatement of Torts from reflecting this most important development in tort causation.

In addition to the evidence-jurisdiction objections, defendants’ representatives decided that, with the revisions reflecting the NAS meeting, they did not like Comment c. Those who supported Comment c at the 2002 meeting advocated removing the revised version from the Restatement the following year.

The irony should not be missed. Defendants’ lawyers like to tout science and its rigorous methodology as essential to taming the evils of junk science. Yet modifications to Comment c after the NAS meeting with scientists resulted in those on the defense side rejecting Comment c and seeking to remove it from the Third Restatement. That scientific reasoning requires subjective judgments about appropriate inferences based on weight of the evidence rather than purely objective deductive conclusions put a different, if more accurate, face on science and its methodology that defendants did not want acknowledged in a restatement.

What seems to have concerned those objecting to Comment c most is that it suggested that the burden on plaintiffs to produce sufficient evidence should take into account the difficulties of obtaining and presenting scientific evidence on a causal proposition. Thus, across the board threshold


41. Sufficiency of evidence is regularly addressed in the substantive Restatements when addressing the proper role of judge and jury. See, e.g., RESTATEMENT (SECOND) OF TORTS §§ 328A–328B (1965).

42. See PROCEEDINGS 2003, supra note 18, at 30–31. Technically, they sought to move Comment c from commentary and relegate it to Reporters’ Notes, which are not reviewed, approved, or endorsed by the Institute. Rather, Reporters’ Notes simply reflect reasoning, explanation, or supporting references by the Reporters.


44. See PROCEEDINGS 2003, supra note 18, at 32–33.
requirements, such as positive epidemiologic studies, should not be imposed.

At the 2005 annual meeting, when the chapter on causation containing Comment c was before the membership for final approval, defense interests once again sought to remove Comment c from the Restatement. A motion and memorandum in support accused Comment c of not stating the law, but rejecting and reinventing it; reflecting a “decided bias” and; misstating scientific principles that the courts, by contrast, had gotten right. Concerns at the meeting by those who opposed inclusion of Comment c ranged from “this Comment simply doesn’t state the law,” to that the Comment took “an intemperate and dismissive view” about courts like the Ninth Circuit in the remand of Daubert. The most entertaining statement was that Comment c “embrace[d] a relativistic sliding-scale approach to proof of causation,” so, with apologies to Stephen Stills, “when you’re not with the evidence you love, love the evidence you’re with.”

45. PROCEEDINGS 2005, supra note 9, at 61.
46. See Memorandum in Support of Motion to Amend Restatement of Torts: Liability for Physical Harm Section 28, Comment c (2005) (on file with author).
47. PROCEEDINGS 2005, supra note 9, at 61.
48. Id. at 62.
49. Id. In objecting to the law adjusting in light of a scarcity of evidence, these objectors ignore a long history of such adjustments. Thus, we should recall that probabilistic evidence such as epidemiology is one that courts historically have rebuffed. The classic case reflecting this attitude is Smith v. Rapid Transit, Inc., 58 N.E.2d 754 (Mass. 1945); see also Guenther v. Armstrong Rubber Co., 406 F.2d 1315, 1318 (3d Cir. 1969). The acceptance of epidemiologic evidence as proof of causation involved a substantial change in judicial attitudes about probabilistic evidence that was presaged in David Rosenberg’s article, The Causal Connection in Mass Exposure Cases: A “Public Law” Vision of the Tort System, 97 HARV. L. REV. 849 (1984). That adjustment occurred, I suggest, because better evidence of agent-disease causation is unavailable.

The acceptance of epidemiology is just one of many examples of the legal system’s adaptation to difficulties of proof. Similar to probabilistic evidence is res ipsa loquitur, which sanctions use of generalized evidence for proof that, were good particularistic evidence available, we would not accept. Indeed, that is why a defendant can defeat res ipsa loquitur by demonstrating that, in the particular instance involved, there was no negligence. The courts that accepted market share liability did so because evidence of which drug manufacturer provided the DES that the plaintiff’s mother took was unavailable. Justice Traynor’s famous concurrence in Escola v. Coca Cola Bottling Co., 150 P.2d 436, 440 (Cal. 1944), advocated the adoption of strict products liability, in part, because of the difficulty plaintiffs had in proving a manufacturer’s negligence. The “substantially certain to result” alternative standard for intent in tort law is justified, in part, by the difficulties of proving a defendant’s purpose. Another area in which the evidentiary standard reflects the difficulties of proof is the sufficiency standard for the magnitude of damages caused by a defendant’s wrongdoing. There the law only expects a
After a back and forth discussion between supporters of the motion and those who opposed it, the motion to remove Comment c was defeated. Comment c took its place in the Third Restatement in the volume that was published in 2010.

However, there was very little judicial attention to Comment c, which we thought would be of enormous benefit to courts. Other provisions in that Restatement were being utilized and adopted by courts. But Comment c languished until the Milward case was decided.

After all of the apologies, blood, sweat, and tears involved in escorting Comment c to the finish line, Milward was a very satisfying opinion to read. Among the propositions it embraced, often employing Comment c in support, were:

- Recognizing that judgment and interpretation are required in assessments of causation.
- Endorsing explicitly and taking seriously weight of the evidence methodology, against the great majority of federal plaintiffs to provide proof with “as much certainty as the nature of the tort and the circumstances permit.” Restatement (Second) of Torts § 912 (1979); see also Restatement (Third) of Torts: Apportionment of Liability § 26 cmt. h (2000) (advocating relaxing the burden of production when good evidence of the extent of harm caused by multiple parties is unavailable); David W. Robertson, The Common Sense of Cause in Fact, 75 Tex. L. Rev. 1765, 1795 (1997) (stating that when apportioning on the basis of causation between a preexisting condition and a tortfeasor’s conduct, juries are instructed “to do the best they can to make the distinction” between the preexisting condition and the “aggravation” brought about by the defendant’s negligent conduct”). Numerous other instances of the law accommodating deficiencies in available evidence might be added to this list.

51. See Restatement (Third) of Torts: Liability for Physical and Emotional Harm § 28 cmt. c (2010).
53. Id. at 17–18 (citing the testimony of plaintiff’s expert Carl Cranor, who had written about the use of weight of the evidence methodology by scientists and scientific organizations such as the International Agency for Research on Cancer and the Institute of Medicine of the NAS to make judgments about suspected carcinogens and other toxic agents); see also Carl F. Cranor, Toxic Torts: Science, Law, and the Possibility of Justice 136–37 (2006); General Elec. Co. v. Joiner, 522 U.S. 136, 152–54 (Justice Stevens dissenting on the grounds that the weight of the evidence methodology employed by plaintiff’s expert was a recognized and reliable form of scientific inquiry). Other commentators have joined in the endorsement:

It is not uncommon for causal relationships to be inferred by the convergence of information from various domains at some remove from the target issue, where the product of no single domain could be
courts that had, since Joiner, employed a Balkanized approach to assessing different pieces of evidence bearing on causation.  

- Appreciating that because no algorithm exists to constrain the inferential process, scientists may reasonably reach contrary conclusions.
- Not only stating, but taking seriously, the proposition that epidemiology demonstrating the connection between plaintiff’s disease and defendant’s harm is not required for an expert to testify on causation. Many courts had stated that idea, but very few had found non-epidemiologic evidence that satisfied them.

III. Milward’s Potential

So, why am I not joining those participating in this Symposium who are praising Milward as a fresh breath of Daubert air, rejecting the corpuscular approach of most federal courts, as Joe Sanders documents in his paper? Instead I am here to express pessimism about Milward’s potential for more intelligent, coherent, and rational causation assessments in future toxic tort litigation.

said to be a reliable indicator of causation by itself. This is not surprising. It is the normal way of circumstantial evidence, building walls by bricks in ordinary trials. When there are interlocking and mutually corroborating results from a variety of domains and studies that individually are all subject to plausible external validity objections, it would seem that exclusion based on external validity grounds ought to be approached with caution and an attempt at sophistication.


54. See Sanders, supra note 7, at 16 n.92.
55. Milward, 639 F.3d at 18 (citing and quoting Comment c).
56. Id. at 24.
57. See Murphy Horne & Hannah Davis, The Uphill Battle of the Toxic Tort Plaintiff: Assessing the Admissibility of Epidemiology in Toxic Tort Litigation (Spring 2011) (unpublished student work, Wake Forest University School of Law) (on file with author) (finding that in eighty-five percent of reported cases in which plaintiffs did not have supportive epidemiologic evidence, expert testimony on causation was ruled inadmissible).
58. See Sanders, supra note 7, at 16 n.92.
There are two reasons, one borne of the difference between the scientific process (at, say, a convention of toxicologists, epidemiologists, or statisticians) on the one hand and the adversary litigation context that Milward addressed. The other reason stems from an episode during the early Bendectin litigation that I recalled, quite vividly, when thinking about Milward and its implications for the future.

First, a preliminary observation, which is important to what follows. We should appreciate that cases that actually get to appellate courts, such as Milward, are not the same as other toxic causation issues. Consider a couple of toxic causation issues: 1) Does thalidomide cause birth defects, mostly horrific birth defects? 2) Does asbestos cause mesothelioma?

The answer to both questions is yes, and whether we were at a Teratology, Occupational Disease, or Oncology conference, all respectable scientists would concur on the answers to those questions. We would find very little variance, if any, in the opinions of these scientists based on the clarity and probity of the scientific evidence addressing the matter. The point I want to make here is that the evidence for those propositions resembles Figure 1:

---

Thus, in Figure 1, all of the evidence converges on one answer, and there is very little or nothing that might support a different inference. By contrast, consider an evidentiary distribution like that in Figure 2. This depiction is a better model for cases like *Milward* and the causal question confronted there and in many toxic tort cases.

**The Reality of Toxic Causation Litigation**

![Diagram](image)

**FIGURE 2**

To take a specific example, consider the drug Parlodel, which was prescribed to suppress lactation in postpartum women.\(^{60}\) There was a considerable body of litigation over Parlodel by women who suffered mostly strokes while taking the drug.\(^{61}\) The frequency of strokes in postpartum women is greater than it is

---


among non-postpartum women similarly situated. Nevertheless, even among this group the incidence is exceedingly rare.

Unlike strong toxicants like thalidomide or asbestos, if Parlodel does cause strokes it does so at a quite modest rate, making epidemiologic investigation into this question extremely difficult. Existing evidence includes a study conducted by Kenneth Rothman, a prominent epidemiologist. Rothman employed a case-control methodology, one that has the best potential to find an effect when the incidence of the investigated disease is very small. However, because strokes in the period after pregnancy are so rare (on the order of 5 per 100,000 pregnancies), Rothman’s study could only identify ten cases to examine.

Rothman and his coauthors concluded—although they found a modest association—that such a small case series would not permit a statistically meaningful inference of causation to be drawn, as also was the case in Milward in which the plaintiff suffered from a rare subtype of acute myelogenous leukemia. Three other epidemiologic studies were inconclusive regarding Parlodel because of methodological limitations. Plaintiffs’ experts found adverse drug event reports that they claimed

62. See generally Steven J. Kittner et al., Pregnancy and the Risk of Stroke, 335 NEW ENG. J. MED. 768 (1996) (discussing the relative risk of stroke for pregnant women compared with nonpregnant women).

63. See Dunn, 275 F. Supp. 2d at 680.

64. The most readily determined agent-disease causal relationship through epidemiologic inquiry is one in which the disease occurs relatively frequently; yet, it is infrequently caused by anything other than the suspected agent. The extreme are pathognomonic causes or signature diseases where the agent is the only cause of the disease and the existence of the disease is a “signature” of the agent’s involvement. Conversely, the most difficult agent-disease relationship to investigate is a disease that is rare and an agent that, if it has any causal role, only causes a small proportion of the disease.


66. Id.

67. Id.


supported causation.\textsuperscript{70} They also relied on biological mechanism evidence derived from chemical structure similarity—the active ingredient in Parlodel is within a class of drugs that are capable of blood vessel constriction in some patients, and vasoconstriction is at least one of the causes of strokes that result from insufficient blood flow.\textsuperscript{71} 

In\textit{ vivo} animal studies were also proffered, although their primary value was confirming aspects of the biological mechanism evidence—Parlodel can lead to vasoconstriction, which can lead to strokes.\textsuperscript{72}

Carl Cranor reported on the outcome of these cases: “[T]he courts disagreed on whether to admit essentially the same kind of evidence and same kinds of experts. Moreover, both district and appellate courts are disagreeing about whether to admit expert testimony and the supporting scientific evidence that Parlodel has the potential for strokes and heart attacks.”\textsuperscript{73}

Parlodel exemplifies the kinds of cases in which an expert consensus will not exist because of the paucity and, in the case of Parlodel, the balance of what evidence there is. In this environment, scientists and expert witnesses will be influenced by their experiences and world views, i.e., their personal biases, and a second and very powerful bias, expert adversarial bias.\textsuperscript{74}

\textsuperscript{70} Siharath, 131 F. Supp. 2d at 1356–57.

\textsuperscript{71} Compare Siharath, 131 F.Supp 2d at 1347 (biological-mechanism evidence insufficient to permit plaintiff’s expert witnesses to testify to general causation), with Tobin v. Astra Pharm. Prod., Inc., 993 F.2d 528 (6th Cir. 1993) (plaintiff’s expert relied predominantly on pathogenic evidence), and Globetti v. Sandoz Pharm., Corp., 111 F. Supp. 2d 1174 (N.D. Ala. 2000) (crediting expert witnesses who reasoned that because Parlodel is a vasoconstrictive agent it has the capacity to cause spasms that result in a heart attack).

\textsuperscript{72} Siharath, 131 F. Supp. 2d at 1355.

\textsuperscript{73} CRANOR, supra note 53, at 25.

\textsuperscript{74} This may overstate the representativeness of the Parlodel litigation. Even excluding adversarial expert bias—see infra text accompanying notes 75–77—in one case, three court-appointed experts disagreed on whether plaintiff’s experts’ methodology or technique was scientifically reliable. Two of the experts concluded that the methodology was not scientifically reliable while the third opined that it was. See Soldo v. Sandoz Pharm. Corp., 244 F. Supp. 2d 434, 503–04 (W.D. Pa. 2003); Soldo v. Sandoz Pharm. Corp., No. 98-1712, 2003 WL 22005007, at *1 (W.D. Pa. Jan. 16, 2002) (order that reports of court-appointed experts David Flockhart, William J. Powers, and David Savitz be filed).
David Bernstein helpfully taxonomizes these adversarial biases. He identifies three different aspects to expert adversarial biases: 1) conscious bias; 2) subconscious bias; and 3) selection bias.

Bernstein defines those who exhibit conscious bias as “hired guns” who “adapt their opinions to the needs of the attorney who hires them.” In contrast with Bernstein, who sees adversarial experts as willing to say whatever their employer wants, I tend to think of this bias as more subtle than Bernstein’s sledgehammer characterization: “Testifying experts shade their testimony, eliminate qualifications, state their conclusions more forcefully, or otherwise modify their prior, uninfluenced views in the crucible of an adversarial proceeding and its impact on those experts. The result is that expert witnesses frequently present conflicting testimony to the factfinder.”

With regard to unconscious bias, Bernstein provides a quotation from a nineteenth century case: “Undoubtedly there is

---


But the practice [of using expert witnesses] under the present method has for years exhibited shortcomings which are lamentable. . . . The principal feature of the breakdown seems to be the distrust of the expert witness as one whose testimony is shaped by his bias for the party calling him. That bias itself is due, partly to the special fee which has been paid or promised him, and partly to his prior consultation with the party and his self-committal to a particular view. His candid scientific opinion thus has had no fair opportunity of expression, or even of formation, swerved as he is by this partisan committal.

JOHN HENRY WIGMORE, A TREATISE ON THE ANGLO-AMERICAN SYSTEM OF EVIDENCE IN TRIALS AT COMMON LAW § 563 n.1 (2d ed. 1923).

76. Bernstein, supra note 75, at 454.

77. Id. at 454–55.

78. Tim Cramm et al., Ascertaining Customary Care in Malpractice Cases: Asking Those Who Know, 37 WAKE FOREST L. REV. 699, 719–20 (2002). John Langbein reports, based on personal experience, in very similar fashion:

I sometimes serve as an expert in trust and pension cases, and I have experienced the subtle pressures to join the team—to shade one’s views, to conceal doubt, to overstate nuance, to downplay weak aspects of the case that one has been hired to bolster. Nobody likes to disappoint a patron; and beyond this psychological pressure is the financial inducement. Money changes hands upon the rendering of expertise, but the expert can run his meter only so long as his patron litigator likes the tune.

a natural bias to do something serviceable for those who employ you and adequately remunerate you.” 79 Yes, I think that is true; indeed, I observe it every year when I read final examination answers by my students who, put in the role of counsel for a party and asked to write a legal memorandum, end up emphasizing the strengths and minimizing the weaknesses of their client’s position, 80 sometimes comically so. 81

Rather than categorizing this as unconscious bias, I would characterize it as identification or affiliation bias—we tend to identify with those who employ and pay us—and we tend to take a course, to the extent that there is ambiguity or uncertainty, in favor of the principal. There is reasonably strong empirical evidence, both survey and observational, and across a wide swath of expert disciplines, that supports the existence of affiliation litigation bias. 82 Those biases are most evident when judgment is

79. Bernstein, supra note 75, at 455–56 (quoting Abinger v. Ashton, 17 L.R.Eq. 358, 374 (Ch. 1873)).

80. My observational evidence of this “affiliation bias” displayed by law students reveals that there is more than simply financial incentives involved in producing this bias.

81. Linda Babcock and George Lowenstein found such a bias in an experimental study that asked students to negotiate a settlement in a tort case. The student-participants were assigned to negotiate on behalf of plaintiff or defendant and were provided identical information about the case. Before beginning negotiations, the participants were asked to guess what a judge would award in the case and provide what they considered a “fair” settlement of the case. Assessments on both of these metrics were skewed by the affiliation assigned to the participant. Linda Babcock & George Lowenstein, Explaining Bargaining Impasse: The Role of Self-Serving Biases, 11 J. ECON. PERSP. 109 (1997).

82. Researchers have examined inter-rater consistency using a standardized test to evaluate the risk of future violence in civil commitment proceedings and found lower inter-rater agreement when the evaluators were involved in adversarial proceedings than in other research or clinical contexts. These researchers suggest that adversarial allegiance is likely responsible for some of that variance. See Daniel C. Murrie et al., Does Interrater (Dis)agreement on Psychopathy Checklist Scores in Sexually Violent Predator Trials Suggest Partisan Allegiance in Forensic Evaluations?, 32 LAW & HUM. BEHAV. 352 (2008); Marcus T. Boccaccini et al., Do Some Evaluators Report Consistently Higher or Lower PCL–R Scores than Others? Findings from a Statewide Sample of Sexually Violent Predator Evaluations, 14 PSYCHOL. PUB. POLY & L. 262 (2008). Studies of professional auditors, business students, and undergraduate students assessing a company’s value and its compliance with generally accepted accounting principles reveal a consistent bias in favor of the entity with which the participant was assigned. See Max H. Bazerman et al., Why Good Accountants Do Bad Audits, 80 HARV. BUS. REV. 97 (2002); see also Lawrence A. Ponemon, The Objectivity of Accountants’ Litigation Support Judgments, 70 ACCT. REV. 467 (1995) (finding that accountants providing litigation support services render different damage valuations depending on whether they were in role of working for plaintiff or defendant); Blair H. Sheppard & Neil Vidmar, Adversary Pretrial Procedures and Testimonial Evidence: Effects of Lawyer’s Role and Machiavellianism, 39 J. PERSONALITY & SOC. PSYCHOL. 320 (1980).
required—precisely the task that weight of the evidence methodology requires.

The third aspect of bias that Professor Bernstein identifies is selection bias, a phenomenon that Joe Sanders wrote about in a superb piece that explained that the fault in the Bendectin litigation was not in the Jury (or the Gods) but in the structure of the legal system. This structure presents to juries two contending experts without the slightest hint that one may come from among a small group of fringe scientists who still believe that global warming is a conspiracy of alternative energy companies, while the other represents the vast majority of global warming scientists.

Selection bias affords attorneys the opportunity to capture inherent biases and employ them to their respective client’s advantage. There are some accomplished, credentialed experts who have never seen a birth defect that was the result of environmental causes. And, conversely, there are some who believe that the causes of birth defects are environmental, even if we have not yet identified the specific environmental agent. More realistically, some experts lean in one of those directions and others in the other. And that is how attorneys on both sides sort out the experts that they hire.

The point is that with the variability in evidence that the Milward and Parlodel cases present, we will have the same dueling adversarial experts, one testifying that using the weight of the evidence methodology leads her to the conclusion that plaintiff’s

(finding in student experiment that witnesses interviewed by an adversarial lawyer skewed their subsequent trial testimony in favor of the lawyer’s client).


85. I should add that those biases would be displayed with regard to the substantial proportion of birth defects that are the result of unknown or uncertain, rather than established, causes. See BIRTH DEFECTS ENCYCLOPEDIA, at ix (Mary Louise Byuse ed., 1990) (stating that in most instances of birth defects the cause is unknown); Robert L. Brent, Environmental Causes of Human Congenital Malformations: The Pediatrician’s Role in Dealing With These Complex Clinical Problems Caused by a Multiplicity of Environmental and Genetic Factors, 113 Pediatrics 957, 958 (2004); Robert L. Brent et al., Clinical Teratology, 5 Curr. Opin. Pediatrics 201 (1993) (noting that sixty-five to seventy-five percent of birth defects are of unknown origin); Robert L. Brent, The Complexities of Solving the Problem of Human Malformations, 13 Clinics in Perinatology 491, 492–93 (1986) (citing various estimates of the proportion of birth defects due to unknown causes).

disease was caused by defendant's toxic agent, and the other testifying to all of the flaws in the evidence relied on by the other expert and why it is not proper scientific methodology.

This is far from the measured assessments that might occur among scientists at an International Agency for Research on Cancer (“IARC”) meeting87 to review the carcinogenicity of an industrial chemical, or an Advisory Committee formed by the FDA to assess evidence of adverse effects of a recently approved new drug. Not only is the selection and preparation of scientists for such proceedings dramatically different from litigation, the process of consensus advisory committees and the process of direct and cross examination of expert witnesses in a courtroom are equally disparate.88

There are two critical sentences in Comment c that expressed just this concern. After explaining that scientists recognize that inferences about causation require judgment and interpretation, which justifies their assessment that reasonable scientists can take different views of a causal question,89 Comment c(1) states: “These scientists' views reflect their scientific experience outside the courtroom. They may have different views about specific instances of conflicting scientific testimony in a courtroom.”90 The idea behind those sentences is precisely the problem of adversarial bias addressed above.

87. In contrast with the use of adversarial experts described above, the IARC explanation of the role of a Working Group in performing a carcinogenicity assessment is revealing:

The Working Group is responsible for the critical reviews and evaluations that are developed during the meeting. The tasks of Working Group Members are: (i) to ascertain that all appropriate data have been collected; (ii) to select the data relevant for the evaluation on the basis of scientific merit; (iii) to prepare accurate summaries of the data to enable the reader to follow the reasoning of the Working Group; (iv) to evaluate the results of epidemiological and experimental studies on cancer; (v) to evaluate data relevant to the understanding of mechanisms of carcinogenesis; and (vi) to make an overall evaluation of the carcinogenicity of the exposure to humans.


88. See Gross, supra note 86, at 1158–76 (describing direct and cross-examination of expert witnesses).

89. RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYSICAL AND EMOTIONAL HARM § 28(a) cmt. c(1) (2010).

90. Id.
Let me conclude with a telling anecdote from the Bendectin litigation—one that so far as I know has never been put in print, despite the academic attention to that mass tort. In the first Bendectin trial, in January 1980, one of the two plaintiff’s experts was a physician named Alan Done, a pediatrician and pharmacologist at Wayne State University Medical School.91

When Done testified at that trial, there was but one epidemiologic study of Bendectin and its teratogenicity.92 It was a study performed in-house by the manufacturer (the Wm. S. Merrell Co.) that was so poorly performed that Merrell ultimately ceased efforts to defend its validity, instead attributing it to being a product of earlier times and dated methods.93

Thus, at the trial of this first Bendectin case, Done focused on adverse event reports for Bendectin that had a high percentage—over sixty percent—identifying birth defects as the adverse event and chemical structural analyses that revealed that the antihistamine ingredient in Bendectin had a similar chemical structure to other antihistamines that had been identified as teratogenic,94 and a toxicologic in vivo study performed by a young Merrell researcher that revealed disquieting malformations in the kits of rabbits who were fed Bendectin.95 All of these were less-than-gold-standard pieces of evidence, but each, perhaps with the exception of the frequency of birth defects in adverse event reports,96 provided relevant evidence pointing to teratogenicity.

92. See id. at 330; see also id. at 104–05 (describing how there had been three other studies of drugs and their teratogenic potential, but none were specific to Bendectin and all, because of their methodology and flaws, provided quite weak evidence to exonerate Bendectin as a teratogen).
93. GREEN, supra note 91, at 330.
95. GREEN, supra note 91, at 128.
96. There are several difficulties in evaluating whether the frequency of birth defects reflected in adverse event reports has any probity: 1) Because Bendectin’s FDA approved use was for morning sickness, only pregnant women used the drug, so the proportion of birth defects in that cohort would, even if Bendectin was not a teratogen, be higher than for other drugs; 2) Birth defects are a prominent harm that often spur a search for the cause and therefore are more likely to be reported than other adverse events; 3) Bendectin’s association with other adverse events is unknown—if it has a low adverse event rate in relation to other drugs, the proportion of birth defects reported would be greater.
But by the time of the next Bendectin trial, Oxendine v. Merrell Dow Pharmaceuticals, Inc., three years later, more exonerative epidemiologic evidence had been published. That epidemiologic research, spurred by the Bendectin litigation, was surely inadequate to prove that Bendectin was absolutely safe, but it did provide a substantial barrier to plaintiffs’ being able to make out a sufficient case on causation.

Thus, Dr. Done developed and testified to his “mosaic theory” of Bendectin’s teratogenicity, employing (1) structure activity considerations, (2) in vitro animal studies, (3) in vivo animal studies, and (4) his interpretation of the available human epidemiological data on Bendectin’s relationship to birth defects.

The jury found for the plaintiff in Oxendine, awarding $750,000 to a child who was born with a shortened right forearm, two missing fingers on her right hand, and webbing between the remaining three. On post-trial motions, the court granted judgment as a matter of law for the defendant, concluding that plaintiffs’ evidence of causation was insufficient to meet the burden of production. The Court of Appeals for the District of Columbia reversed, commenting on Dr. Done’s testimony:

Like the pieces of a mosaic, the individual studies showed little or nothing when viewed separately from one another, but they combined to produce a whole that was greater than the sum of its parts: a foundation for Dr. Done’s opinion that Bendectin caused appellant’s birth defects. The evidence also established that Dr. Done’s methodology was generally accepted in the field of teratology, and his qualifications as an expert have not been challenged.

98. See GREEN, supra note 91, at 173.
99. See id.
101. Id. at 1103.
102. Id. at 1110.
103. Id.
As a later court, remarking on the court of appeals opinion, put it: “The court of appeals was impressed by the careful and thorough nature of Dr. Done’s testimony, in particular by his admissions that the first three types of evidence, while probative, could not definitively establish that Bendectin is a teratogen, and that no single epidemiological study demonstrated Bendectin’s teratogenicity.”

Sound like weight of the evidence? It sure does to me. A footnote on Dr. Done: in the midst of the Bendectin litigation, he was dismissed from his academic position because of charges that he had become a professional expert witness and neglected his academic duties in a way that brought “discredit” to his employer, the medical school at Wayne State University. Those facts emerged in a later proceeding in Oxendine, in which Merrell sought once again to overturn the plaintiff’s verdict. The trial judge’s opinion, overturning the verdict, is scathing:

The court found that Dr. Done lied in several respects. . . . The court finds that his testimony was so deliberately false that all his testimony on behalf of plaintiff is suspect. His lies went so much toward enhancing his status as a witness that he reeks of the hired gun who will say anything that money can buy so long as it is glibly consistent with his prior testimony in other cases. In a proverbial spiral his professional witness status led him to shirk his duties at the Wayne State Medical School. That got him fired (gently, by a forced resignation). The true circumstances of that resignation detracted from his professional witness status, and so he covered it up with lies to maintain his purported status.

One final observation—the source of Dr. Done’s mosaic theory. While interviewing Barry Nace, a talented and pugnacious plaintiff’s lawyer in the District of Columbia, who tried Oxendine, I inquired about Dr. Done’s mosaic theory and its development to

105. GREEN, supra note 91, at 280.
107. GREEN, supra note 91, at 281.
counter the growing body of exonerative epidemiology. Nace looked at me, smiled, and said something like, “Damn brilliant, and I was the one who thought of it and fed it to Alan [Done].”

I do not think that the Milward case means that Dr. Done could employ his “mosaic” theory (or rename it “weight of the evidence”) to testify today that Bendectin caused a child’s birth defect. Courts are going to have to police weight of the evidence in a similar way that they have been policing expert witness testimony in toxic cases since Daubert, and Milward will not ease that burden to any significant degree in my view.

The right way for courts to do that is, as Joe Sanders and I have argued elsewhere, for the judge to examine the underlying scientific evidence and, with the assistance of the experts in understanding it, determine if that evidence is sufficient to permit a reasonable inference of general and specific causation or if, on the other hand, impermissible speculation is the only way to reach a conclusion of causation. That is a wavering and unclear line that requires careful consideration of the evidentiary record, judgment, and interpretation, and for which there is no shortcut weight of the evidence methodology that can be delegated to adversarial expert witnesses. But it is also the same essential enterprise that courts have long been engaged with in ordinary tort cases when a defendant challenges the evidence proffered by

---

108. Indeed, the Milward court addressed the situation in which there is a substantial body of exonerative evidence, stating: “To be clear, this is not a situation in which the available epidemiological studies found that there is no causal link, or even one in which no cases of APL were found among benzene-exposed workers.” Milward v. Acuity Specialty Prods. Grp., Inc., 639 F.3d 11, 17 (1st Cir. 2011), cert. denied sub nom. U.S. Steel Corp. v. Milward, 132 S. Ct. 1002 (2012). The Milward opinion contains other caveats and statements that no doubt will be the basis for limiting it in First Circuit cases in the future. See, e.g., id. at 22–23 (“Dr. Smith’s opinion was based on . . . substantial evidence that he carefully explained.”); id. at 21 n.15 (noting that Dr. Smith had not ignored but had addressed a study that contradicted his testimony that APL and AML share a common biological mechanism and explained why he discounted it); id. at 24 (“[T]his is a case in which there is a lack of statistically significant epidemiological evidence, and in which the rarity of APL and difficulties of data collection in the United States make it very difficult to perform an epidemiological study of the causes of APL that would yield statistically significant results.”); id. (“[T]his is not a situation in which the available epidemiological studies found that there is no causal link, or even one in which no cases of APL were found among benzene-exposed workers.”).

109. See also Sanders, supra note 7, at 15 n.86.

110. See Green & Sanders, supra note 40.
plaintiff as insufficient to permit a jury finding that causation exists. 111

111. See Restatement (Third) of Torts: Liab. for Physical and Emotional Harm §28 cmt. b (2010). Thus, I am more sanguine than Steve Gold about judges having to accomplish this task. Steve Gold, A Fitting Vision of Science for the Courtroom, 3 Wake Forest J.L. & Pol’y 1, 7 n.38 (2013).