A preliminary history of epidemiological evidence in the twentieth-century American Courtroom

Tal Golan

This chapter follows the short but brilliant career of epidemiology in the American courtroom. Until the 1970s, epidemiological evidence could hardly be found in the legal system. By the 1980s, it was already announced “the best (if not the sole) available evidence in mass exposure cases,” and by the start of 1990s, judges were dismissing cases for not supporting themselves with solid epidemiological evidence. Epidemiology, I show below, owed much of this prosperity to the equally meteoric career of mass tort litigation – a late-modern American species of litigation involving crowds of plaintiffs, all claiming to be harmed by the same exposure or mass-marketed product. Dangerous drugs, industrial defects, environmental pollutants, radiation exposure and other technological breakdowns -- all have become the subject of prolonged mass tort litigation with ever-escalating financial stakes. Questions about risk and causation were central to a great majority of these cases, and in the absence of direct proof of cause and effect the courts increasingly turned to statistical evidence to resolve these questions.

The mutually-constitutive rise of mass tort litigation and statistical science, I suggest below, set the tone of the relations between law and science in the late-20th century

---

1 Michael Green, The Impact of Daubert on Statistically-Based Evidence in the Courts, Proceedings of the American Statistical Association, Statistics in Epidemiology Section (1998), 35
adversarial courtroom. Early in the 20th century, tort, a branch of private law that deals with personal injury claims, still prided itself on its tradition of individualized approach. Its clients were willful and rightful citizens whose causal agency could not be subsumed mechanically, without the careful exercise of human judgment on a case-by-case basis.\footnote{Morton J. Horwitz, The doctrine of objective causation, in D. Kairys ed. The Politics of Law (New York: Pantheon Books, 1992) 201-213.} But as the twentieth century progressed, tort law became less private and more public, and by the end of the century the ‘statistical victim’ became tort’s biggest client and epidemiology its favorite science.\footnote{Sheila Jasanoff, Science and the Statistical Victim: Modernizing Knowledge in Breast Implant Litigation, 32 SOC. STUD. SCI., 37, 38-40 (2002).} With the new client came new practices: individual care gave way to economy of scale, and direct testimony to epidemiological evidence. These were uneasy changes for tort law and they presented the legal mind with a host of difficult problems regarding the differences between statistical correlation and legal causation, the circumstances in which we could pass from one to the other, and how and by whom should these be decided.

This set of problems played an important role in shaping the late 20th century relations between law and science. Having never before addressed the legal processing of scientific evidence, the US Supreme Court found it necessary to visit the topic on three separate occasions during the 1990s, all of them tort cases.\footnote{Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 US 579. (1993); General Electric Co v Joiner, 522 US 136 (1997); Kumho Tire Company, Ltd. v. Carmichael, 526 US 137. (1999).} Known as the \textit{Daubert Trilogy}, these three Supreme Court opinions announced the arrival of a new era in the relations between law and science. The traditional legal deference to scientific expertise was over. The trial judge, who had long remained passive in the play of science in the adversarial courtroom, was charged with a new responsibility of preventing junk science from entering the courtroom and bamboozling the jury.\footnote{Cf. Symposium, ―Scientific Evidence After the Death of Frye,‖ Cardozo Law Review 15 (1994); B. Black, J. F. Ayala and C. Saffran-Brinks, “Science and Law in the Wake of Daubert: A New Search for Scientific Knowledge.” Texas Law Review 72 (1994), 715-802; Angell, Science on Trial, 127.} This new role of the trial judge as a gatekeeper of true science, I suggest below, was co-related with the new role of the statistical expert as the gatekeeper of true causes in mass tort litigation.
The rise of epidemiology

Modern science has offered public decisionmakers two distinct modes of calculating risks and constructing causality: toxicology, an experimental reductionist science, built on the strength of the laboratory; and epidemiology, an observational statistical science built on the power of big numbers. Earlier in the twentieth century the toxicity of things was checked in the laboratory. One strategy, called in vitro studies, examined the effects of chemical agents on various organic materials ranging from DNA and proteins, to cells, bacteria and even embryos, in attempt to understand the biochemical mechanisms involved. Molecular structural analysis was also called upon to gain clues from structural resemblance to other, better known, chemicals.  

It is a long way, however, from molecules to humans, and other researchers have taken a shortcut by performing in vivo studies. This reduced some difficulties but introduced new ones. Unable to experiment directly with humans, the toxicologists run their studies on other mammals. But even though much is common across the mammalian species, much is also different and scientists were not always sure which is which. In addition, in vivo studies typically involve larger-than-life doses, to shorten the experiment and augment the effects. To make these studies policy relevant, toxicologists must then extrapolate from the short and intense exposure of the tested mammals to a chronic low-level exposure of humans. The extrapolation is dubious but it allowed to work the numbers into a dose-response curve that allows calculating the risks per any given dose and any given period, and most importantly for the setting of exposure standards, with appropriate safety factors to protect the more susceptible subpopulations.

---


10 The choice of a specific extrapolation method is problematic. Linear extrapolation is prevalent but there are sometimes good reasons to apply non-linear extrapolations.

During the 1970s, as environmental regulation took central stage in western polity, the capacity of this laboratory science to provide reasons good enough to legitimize administrative action was closely scrutinized. As the young regulatory agencies began to churn out their safety standards, both industry and civil action groups challenged the science behind the standards--industry in attempt to moderate the standards; civil activists, to step them up.\textsuperscript{12} The ensuing legal battles revealed to all the fragility of the science involved. What had thrived in the temperate climate of the laboratory did not survive the adversarial heat of the courtroom. The notorious non-linearity of physiological systems was mobilized to undermine the extrapolations from high to low doses and from short to long exposures; and the poorly understood interspecies and intrahuman variations were called upon to show that the justification of the standards went beyond scientific and technical competence.\textsuperscript{13} Eager to protect the regulatory regime, the legal system responded by adopting the powerful precautionary doctrine, which admitted the fragility of the science involved but justified the right of the authorities to act upon it, based on the ever-pressing need to regulate potential risks before they turn into actual harms.\textsuperscript{14} The legitimacy of such a regulatory regime, the courts prescribed, resided in its deployment of the best scientific tools available. These tools, the judges also increasingly suggested, may no longer be found in the laboratory but in the arsenal of epidemiology.\textsuperscript{15}

Earlier in the 20\textsuperscript{th} century, epidemiology served public policy as a form of surveillance technology.\textsuperscript{16} Medical attention was focused on infectious diseases -- each caused, it was generally held, by a specific microbiological agent. Fighting infectious diseases was a job


\textsuperscript{13} See Reserve Mining Co. v. Environmental Protection Agency (1975) 514 F.2d 492, and Ethyl Corporation v. Environmental Protection Agency (1976) 541 F.2d 1.,

\textsuperscript{14} Jasanoff, Ref 9. Id.


for the laboratory - to isolate the specific causal organism, study it and devise the best means to fight back.¹⁷ Epidemiology served in this campaign merely by informing of geographical and social patterns of the disease. But by the middle of the twentieth century the balance had begun to shift. The battle against infectious diseases seemed to have been won in the developed world, and public and medical attention increasingly turned to a new pattern of diseases: non-infectious, chronic, with long latency and poorly understood etiology; diseases such as blood pressure, cancer, or heart problems- all of which were previously considered inevitable failures of the aging organism and now began to top the medical charts.¹⁸

Experimental science, with its reductionist logic, made little progress with these so-called ‘diseases of civilization’. They seemed to involve multiple causes and effects; their long latency made experimentation difficult, and their mechanisms kept eluding the researchers. Epidemiology, on the other hand, proved much more flexible. A post-facto observational science that relates exposure to outcome, it did not have to ponder too much over the illusive biological mechanisms involved. Instead, epidemiologists adapted their computational strategies to a distributed, multivariate model of causation that seemed to better fit the nature of these new diseases, where a cause could have many effects and an effect many causes.¹⁹

The power of epidemiology to make causal claims in this new web-like universe of irreducible, chronic health problems was first demonstrated during the late 1950s and early 1960s, when a cluster of British and American epidemiological studies first implicated cholesterol and smoking as significant causal factors for heart disease, and in

the case of smoking, also for lung cancer. Running ahead of experimental research, these studies made no appeal to concrete biological mechanisms. Instead, they introduced a new lexicon that appealed only to what came to be known as ‘risk factors’ – environmental, social and other patterns that are statistically correlated with higher incidence of disease; the more robust the correlation the more certain the association. Nevertheless, or precisely because of it, many medical scientists went up in arm. At stake, they cautioned, was no less than the scientific essence of modern medicine, which was very much rooted in the laboratory. Epidemiology, they pointed out, was not an experimental science. It could neither sufficiently control its data nor test the veracity of its conclusions. Thus, while epidemiology remained useful in generating causal hypotheses, only experimental science could reliably validate them.

Criticism of the newly-fangled epidemiology was by no means limited to die-hard experimentalists. Geneticists faulted epidemiology for focusing attention on environmental effects, while social scientists faulted it for focusing attention on

---


21 See for example the language of the landmark 1964 Surgeon General's Advisory Committee report on smoking: “It should be said at once, however, that no member of this committee used the word ‘cause’ in an absolute sense in the area of this study. Although various disciplines and fields of scientific knowledge were represented among the membership, all members shared a common conception of the multiple etiology of biological processes. No member was so naive as to insist upon mono-etiology in pathological processes or in vital phenomena. All were thoroughly aware of the fact that there are series of events and developments in these fields and that the end results are the net effects of many actions and counteractions.” Smoking and Health, The Report of the Advisory Committee To The Surgeon General of the Public Health Service, U.S. Department of Health, Education, and Welfare, Washington, D.C., 1964, p. 23.

individual factors abstracted of social context. The most damaging critique came from within, from bio-statisticians anxious to protect the integrity of their science and from epidemiologists who were concerned that too much would be claimed for their fledgling science that was just starting to make inroads into medicine. These sophisticated critics were able to point out various methodological difficulties inherent to epidemiological research, from selection biases to confounding variables, all of which further undermined epidemiology’s capacity to establish authoritative causal claims.23

The proponents of the new risk-factors epidemiology responded by appealing to usefulness rather than truthfulness. They pointed out that while a clear experimental demonstration of a concrete causal relation may indeed constitute a higher form of proof, it was nevertheless hard to come by in this new era of chronic diseases. In the absence of such strong proof, they prescribed a diet of epistemological modesty and methodological flexibility. The distributed nature of the problem was to be matched by an equally distributed scientific effort. The epidemiologist’s search for health risks was still to be based on the strength of carefully-constructed statistical studies, but epidemiologists should remain mindful of the limitations of their method and careful to support it with other types of evidence. In the absence of a concrete demonstrable mechanism, the epidemiologists should nevertheless look for a plausible biological explanation. In the absence of direct experimental control, the epidemiologists should support their causal hypothesis by plausible temporal and dose-response curves, and indeed by any other coherent source of evidence. None of these explanatory factors was sufficient or necessary, nor could any of them bring forward indisputable evidence for or against the causal hypothesis researched. Epidemiologists should therefore qualify their confidence with appropriate confidence margins, and single studies should be treated skeptically until their results are verified by other studies, conducted by different persons, in various places, circumstances and times. The combined weight of these studies, the

epidemiologists maintained, was in a growing number of cases the best science could offer public health decision-makers in this new era of latent and irreducible causes and chronic disease.\textsuperscript{24}

Disdained by scientific purists, this pragmatic program of epidemiology was warmly embraced by the expanding regulatory regimes of the late twentieth century. Practical by nature, judges, legislators, administrators, and public health officers were less concerned with the rigorous pursuit of experimental design and more with the pressing businesses of public policy, which often necessitated judgment made with less than perfect information.\textsuperscript{25} They found epidemiology with its quantified logic and its focus on the population as the unit of investigation perfectly placed to provide them with potent tools to estimate the prevalence of otherwise irreducible health problems, investigate their probable sources, identify those groups with elevated risks, and target them with preventive measures.\textsuperscript{26}

The later part of the twentieth century saw therefore the flourishing of the so-called “black-box epidemiology” - a technical, policy-driven epidemiology that shunned biological hypotheses and concentrated on computing the risks facing taxpayers from a myriad of modern conditions.\textsuperscript{27} The parallel growth of medical registries and computer technology allowed for the deployment of increasingly complex statistical techniques in the search for increasingly smaller risks in increasingly larger populations. The

\begin{flushleft}
\end{flushleft}
epidemiologists traded up their mechanical rulers first for punch cards and then for software programs, and got comfortable with the new tools of multivariate correlation and regression, and exotic tests of statistical significance. By the end of the 20th century, the reduction of causes to a distributed network of risk factors had become prevalent and increasingly informed medical research as well as regulatory and legal action. In theory, some continued to insist, this was not a science of causation. In practice, however, it was exactly this - a hunt for causes; if not for science then certainly for administrative and legal action.28

The rise of mass toxic torts

The legal search for toxic causation grew dramatically during the 1970s in two main contexts: judicial review of regulatory action and private tort litigation, and the legal system has responded differently to each context. The courts were willing to put forward the precautionary doctrine to legitimize regulatory action even in the absence of unequivocal science, but they were reluctant to extend the same leniency to the private sphere of litigation.29 Claims of potential risk belonged to the public sphere and required a relatively low level of proof to justify regulative action. Anything more, the judges recognized, would leave but few environmental regulations standing. Claims of actual harm were treated differently, however. These belonged to the private sphere where, the courts insisted, the traditional requirements of tort law still held: the plaintiff must offer a persuasive proof of a concrete and actual harm caused by the defendant. Anything less, the courts held, would be unfair to the defendant, who should not be forced to pay for injuries it did not cause.30

The courts’ insistence on a concrete proof and the difficulties of science to deliver it turned causation into a central problem for the thriving late-modern genus of toxic torts

29 Jasanoff, ref 9., 114-137.
litigation. Tort’s tradition of private, individualized justice cultivated a theory of causality as reductive as that of the science of infectious disease. To exists, a legal cause had to be reduced to a causal agent.\textsuperscript{31} This causal agent was a human being, not a microbe, a fact that added a moral dimension and much complexity to the process of proof. Nevertheless, the plaintiff’s burden of proof, like that of the medical experimentalist, was to single out the causal agent and demonstrate the chain of events that linked the agent’s actions to the plaintiff’s injury. If a specific causal agent could not be uniquely determined; if the plaintiff could show only that the defendant’s action might have caused the harm; or if another indistinguishable potential cause existed, the courts dismiss the claim for the failure to prove specific causation.\textsuperscript{32}

This reductionist model of specific causation has worked quite well in traditional tort cases, such as accidents or assaults. The defendant’s identity and conduct could be verified by direct evidence such as eyewitness testimonies, and the causes for a black eye or a flooded house were understood well enough to allow the courts to decide liability based on whether those causes were controlled by the defendant. This was not the case, however, in a growing range of environmental, work-safety, and product liability cases that came to be known by the end of the 1970s as ‘toxic tort’ cases.\textsuperscript{33} These cases involved injuries of the kind that has frustrated experimental science - chronic, with long latency and poorly understood etiology; injuries that could not be comfortably reduced to a single cause. In the absence of direct or experimental proof of cause and effect the courts increasingly turned in these cases to epidemiological evidence. That was particularly true for the new and emerging phenomenon of mass toxic tort litigation that clustered together large crowds with various case histories, all claiming to be harmed by the same exposure or by the same standardized, mass-marketed product. Here, lawyers and judges, just like legislators, administrators and public health officers, found

\textsuperscript{33} The first case to be termed toxic tort was an early Agent Orange case from 1979. See Robert F. Blomquist, “American Toxic Tort Law: An Historical Background,” 1979-87, Pace Environmental Law Review (1992) 10: 85-173, on 86.
epidemiology’s quantified logic and population-based analysis particularly conducive to their needs.\textsuperscript{34}

All this may serve to explain why there were little signs of epidemiological life in the legal system prior to the rise of mass toxic tort litigations in the late 1970s.\textsuperscript{35} If epidemiological experts made appearance in the courts before the 1970s, it was typically in their capacity as public health officers, to inform judge and jury of certain medical facts such as the ability of tattoo parlors to keep their needles sterile,\textsuperscript{36} or that an open pond of raw sewage could generate an outbreak of hepatitis.\textsuperscript{37} I found only four toxic tort cases in the 1960s that involved epidemiological evidence – two tobacco and two vaccine cases. In the tobacco cases, the plaintiffs tried and failed to establish a causal relation between smoking and cancer via the testimony of epidemiological experts. In both cases the judges dismissed the epidemiological evidence and directed summary verdicts for the tobacco companies.\textsuperscript{38} The two vaccine cases came later, at the close of the decade and involved plaintiffs who contracted polio after receiving a live-virus vaccine and sued the manufacturers. In these cases, the defense lawyers set the up-and-coming epidemiologists against the traditional medical experts who testified on causation for the plaintiffs. Here too, epidemiology failed to make an impression. In both cases the juries preferred the testimony of the attending physicians, who testified that the vaccine was more probably than not the cause of the plaintiffs’ polio, and gave the verdicts to plaintiffs.\textsuperscript{39}

\textbf{The Swine Flu litigation}

\textsuperscript{37} Larsen v. Vill. of. Lava Hot Springs, 88 Idaho 64, 73, 396 P.2d 471, 476 (1964)
\textsuperscript{38} Pritchard v. Liggett & Myers Tobacco Co., 295 F.2d 292 (1961); Lartigue v. R.J. Reynolds Tobacco Co., 317 F.2d 19 (1963);
Vaccine-related cases may have been rare at first, but the awards were significant and the trend unsettling.\(^{40}\) Thus, when the US federal government initiated early in 1976 the largest ever national immunization campaign, in attempt to vaccinate virtually the entire adult population against swine flu, the insurance industry considered the program’s potential liability for vaccine-related injuries to be “enormous and worse, uncertain,” and refused to underwrite it “at virtually any price.”\(^{41}\) The two-hundred millions doses requested, combined with the American itch to sue, hostile juries, and the new liability theories explored by the courts, were simply too much for the insurance industry. They were in business to spread risk, not to take it. “If the public was really endangered,” retorted the president of one insurance company, “the government should take the risk; it certainly could, we couldn’t.”\(^{42}\)

Without proper liability insurance the manufacturers refused to produce the two-hundred millions shots requested. Pressed in time and left with little choice, the federal government reluctantly agreed to take over the liability from the manufacturers and the litigation costs from the insurers.\(^{43}\) On October 1, 1976 the ambitious vaccination campaign started, and the National Communicable Disease Center [CDC] established an equally-ambitious surveillance system that operated twenty-four hours a day and used the best available computer technology to monitor the campaign. By December 1, more the forty millions American had received the flu shot; twice the number ever immunized before for any disease. Alas, except for one possible case in Concordia, Missouri, there was no sign of the swine flu. Meanwhile, the CDC was receiving alarming reports about the appearance among the vaccinated of the Guillain-Barré Syndrome [GBS], a rare but serious auto-immune disorder. Consequently, on December 16, 1976, the federal government announced the suspension of its swine flu program until the GBS risk is evaluated.\(^{44}\)

\(^{40}\) Cf. 1974 - Reyes vs. Wyeth Laboratories, 498 Federal Reporter, 2nd series, 1264.
\(^{42}\) Id., 46.
\(^{43}\) Id.
\(^{44}\) Id. Under the circumstances, this meant virtual termination
When the immunization campaign ended, the legal battle began. By the start of the 1980s, more than 4,000 wine-flu claims have piled up against the US government. To help the Department of Justice prepare its legal strategy, the epidemiologists at the CDC used the data they collected to study the relations between the swine-flu vaccine and GBS. The studies revealed a clear association between vaccination and GBS, provided the onset of the disease was within a few weeks of vaccination. Based on these studies, the Department of Justice decided not to contest claims in which GBS appeared within six weeks of the vaccination. Of the more than 4,000 claims, 2813 were denied, out of which 1604 led to lawsuits. By 1985, 307 of these lawsuits ended in trial and the government won 259 of them, an impressive 84% win rate.

The growing helpfulness of epidemiology in deciding the slippery question of causation in toxic tort cases was further demonstrated in the late 1970s by two other large and highly-publicized mass-toxic tort litigations involving asbestos and the first synthetic hormone, diethylstilbestrol (DES). In the swine-flu litigation epidemiological evidence mainly served to refute causation; but in the asbestos and DES litigations it changed sides and established causation by demonstrating a strong correlation between the exposure and a unique ‘signature’ disease among the exposed. Adenocarcinomas of the vagina


46 The legal expanse paid by the government was approaching the original budget of the entire swine flu program, which was estimated at $134 million. Rheingold & Shoemaker, The Swine Flu Litigation, 8 Litigation 28 (Fall 1982); Reitz. AW. Jr. Federal compensation for vaccination induced injuries. Boston Coll. Environ. Aff L Rev. 1986;13:169-214.

47 Diethylstilbestrol (DES) was widely prescribed during the 1950s and 60s to to pregnant women to prevent miscarriage.

48 Adenocarcinomas of the vagina and uterus were almost unknown among women whose mothers had not taken DES, and Mesothelioma, another form of cancer, was alleged also to be uniquely associated with asbestos exposure. See , , .
and uterus were almost unknown among women whose mothers had not taken DES, and mesothelioma, another form of cancer, was alleged also to be uniquely associated with asbestos exposure. These exclusive relations allowed the plaintiffs to argue that the litigated exposure was responsible for their specific ailment and win decisive legal victories against the manufacturers.\textsuperscript{49} The successes of the asbestos and DES plaintiffs brought a rising tide of toxic tort actions to the courts in the early 1980s. The two largest actions were \textit{Allen v. United States} and \textit{Agent Orange}, and each of them presented fresh challenges to the judicial embrace of epidemiology.

\textit{Allen v. United States}

For twelve years, between 1951 and 1963, the US government detonated more than 100 atomic bombs at test sites above and below the southern Nevada desert. Three decades later, in the early 1980s, civilians who lived in the neighboring regions entered 1192 individual lawsuits against the government, accusing it of negligence and carelessness in carrying out the tests and demanding hundreds of millions of dollars in damages for hundreds of radioactive-related deaths and injuries.\textsuperscript{50} By the 1980s, scientific research, including studies of surviving victims of WWII atomic warfare, had left little doubt that ionizing radiation can indeed cause cancer.\textsuperscript{51} Still, the downwinders, as the plaintiffs came to be known, suffered from all kinds of cancers, many of which could be found also in the general population and could have resulted from causes other than the exposure to radioactive fallout. The downwinders found it therefore extremely difficult, if not impossible, to satisfy the legal demand for a proof of specific causation and persuade the

\textsuperscript{49} In both litigations there was a problem in identifying the particular manufacturer that caused the injury, forcing the courts to allocate the damages according among the manufacturers, according to their market share.


court that their ailments would not have occurred but for the radioactive fallout from the nuclear testing.

Despite the lack of an adequate proof of specific causation, Bruce Jenkins, the federal district judge who tried the litigation, refused to dismiss the case. In a 489-pages massive opinion, Jenkins assembled ample precedents to show that in cases in which the defendant’s conduct was manifestly tortuous but the plaintiff had no means of identifying the specific cause of injury, the courts had taken steps to ease the plaintiff’s burden of proof by shifting some of it to the defendant. Jenkins considered Allen v. United States to be such a case. He found the government negligent not only for failing to provide off-site civilians with adequate warnings and protection from the radioactive fallout, but also for failing to adequately monitor and record off-site exposures, thereby depriving the plaintiffs of information crucial for the proof of causation. In these circumstances, Jenkins reasoned, “[causal] analysis using ‘but – for’ tests in any form falls short of the mark.” Instead, the requirements should reflect both the objective difficulties involved in the proof of causal relation between radiation and non-specific cancers, and the government’s responsibility for encumbering these difficulties. Thus, Jenkins ruled, it was sufficient for the downwinders to present properly-supported epidemiological evidence “from which reasonable men may conclude that it is more probable that the event was caused by the defendant than it was not.” Once the plaintiff had done so, Jenkins prescribed, the burden of proof will shift to the government to produce evidence extricating itself from the tangle of causality.

Allowing each of the 1192 individual plaintiffs to have his or her day in court was a tall order, especially since it was not clear whether an appellate court would not exempt the government from claims for damages caused by policy decisions. Jenkins decided therefore to test the water first by trying a group of twenty-four cases, selected out of the

52 Allen v. United States, 588 F. Supp. 247 (D. Utah 1984), 415. This was a juryless action, conducted under the Federal Tort Claims Act, which authorizes suits for damages against the United States. See 28 U.S.C. 1346(b), 2401(b), 2671-80.
53 Jenkins’ two main precedents were Summers v. Tice, 33 Cal.2d 80, 199 P.2d 1 (1948), and Basko v. Sterling Drug Co., 416 F.2d (1968).
54 Allen v. United States, Ref. 52.
nearly 1200 claims on his docket.\textsuperscript{55} In deciding these ‘bellwether’ cases, Jenkins relied heavily on the epidemiological studies available. In nine of these cases that involved leukemia and thyroid cancer, the numbers demonstrated a significant increased in the incidence of the cancer within the exposed population, and Jenkins held for the plaintiffs. In fourteen cases that involved other cancers and lacked convincing statistical evidence, he ruled against them.\textsuperscript{56}

As expected, In 1987, Jenkins’s ruling was overturned by the 10th Circuit Court of Appeals on the basis that the United States was protected by the legal doctrine of sovereign immunity.\textsuperscript{57} However, the appellate decision did not discuss Jenkins’ innovative decision to rely on epidemiological evidence in establishing factual causation, even when other possible causes could not be excluded, and it remained standing. Still, the success of Jenkins’ strategy depended on the availability of an authoritative body of epidemiological research that could compensate for the lack of direct evidence and allow for a causal determination even in the presence of alternative causes. But this could hardly be expected in many mass tort actions, given the scarcity of even the most basic toxicity data.\textsuperscript{58} How was the court to decide causation then, in the absence of an authoritative scientific advice? This question stood at the center of the largest and most publicized mass toxic tort litigation of the 1980s – the Agent Orange case.

\textbf{Agent Orange}

The Agent Orange action was brought by many thousands of Vietnam veterans who believed they had suffered or might suffer a variety of diseases due to their war-time exposure to Agent Orange – an herbicide widely sprayed on Vietnam’s jungles by the

\textsuperscript{55} Id.,247.
\textsuperscript{56} Id. at 446-47. One case remained unresolved.
\textsuperscript{57} Allen v. United States, 816 F.2d 1417 (10th Cir. 1987)
U.S. military in order to destroy the jungle and the advantages it afforded to the enemy. Agent Orange contained minute quantities of dioxins, a family of highly toxic compounds that the veterans believed were responsible for their health problems, which included cancers, heart attacks, a suppressed immune system, hormonal imbalances, diabetes, menstrual problems, increased hair growth, and weight loss. ⁵⁹

Much was in common between Allen v. United States and Agent Orange. As with ionizing radiation, little doubt existed about the severity of dioxins at high doses but far less was clear about their impact at lower doses. Like in Allen, the specific levels of individual exposure to Agent Orange were unknown and had to be reconstructed from insufficient military records and from personal memories, many years after the fact. Like the downwinders, the Vietnam veterans suffered from a variety of ailments that could be found in the general population and could not be reduced exclusively to dioxin exposure. And like Jenkins, Jack Weinstein, the federal district judge who managed the Agent Orange case, was willing to rely on epidemiological studies alone to establish factual causation. “We are in a different world of proof than that of the archetypical smoking gun,” Weinstein noted. “We must make the best estimates of probability that we can, using the help of experts such as statisticians and our own common sense and experience with the real universe.” ⁶⁰

Unlike Allen v. United States, however, the best estimates of probability in Orange Agent left much in doubt regarding the capacity of Agent Orange to cause the alleged harms. The epidemiological studies undertaken by the federal government and various state agencies failed to demonstrate a statistically-significant increase in the rate of relevant ailments among the veterans and their families. The only alleged injury that was demonstrably correlated with exposure to Agent Orange was chloracne, a disturbing but

⁶⁰ In re: Agent Orange Product Liability Litigation, 597 F. Sup. 749 (1984), section B.3: Possible Solution in Class Action
far from fatal form of acne.\textsuperscript{61} Science failed therefore to find a causal connection between Agent Orange and the veterans’ ailments. This left Judge Weinstein facing the following dilemma: can he find a causal connection where the statisticians failed to find it?

\textbf{Two kinds of error} could be made in the quest for true causes: a false cause could be found (false positive) and a true one could be overlooked (false negatives). Epidemiologists have always been more vigilant about the first kind. To guard against the possibility of claiming associations where they do not exist, the epidemiologists adopted a two-tier defence strategy called the \textit{null hypotheses}. This strategy operates under the presumption that no causal connection exists between the exposure and disease under study, and demands a strong proof to reverse this presumption. The strength of such proof depends not only on the measurement of a high enough risk, but also the statistical assurance that it is not a false association, created by chance alone. Statistics provides such an assurance by calculating the probability of false association, and the epidemiological dogma demands it to be smaller than 5\% (i.e, less than 1 in 20) for the association to be considered statistically significant.

The ‘statistical significance’ standard is far more demanding than the ‘preponderance of the evidence’ or ‘more likely than not’ standard used in civil law. It reflects the cautious attitude of scientists who wish to be 95\% certain that their measurements are not spurious. But such prudence comes with a price. The rates of false positives and negatives are inversely related. Hence, the more you guard against false causes the more you are bound to miss true ones.\textsuperscript{62} Epidemiologists have considered the price well worth paying. So has criminal law, which emphasizes the minimization of false conviction, even at the price of overlooking true crime. But civil law does not share this concern.

\textsuperscript{61} National Academy of Science, Veterans and Agent Orange: Health Effects of Herbicides Used in Vietnam (Washington, DC: National Academy Press, 1993)

Unlike science or criminal law, it has no preference for either false positives or negatives. It only cares for the preponderance of the evidence.

Both Jennings and Weinstein were aware of this incommensurability between epidemiology and civil law. But they differed in their reactions. Jenkins noted that the statisticians’ 95% probability requirement for significance was arbitrary and stringent, and cautioned his colleagues not to constrained themselves by simplistic models of causal probability imposed upon the judicial ‘preponderance of the evidence’ standard. “Like statistical significance,” he wrote, “mathematical probability aids in resolving the complex questions of causation raised by this lawsuit, but is not itself the answer to those questions.” Judge Weinstein, on the other hand, was far less concerned with the strictness of the epidemiology. A scholar of evidence, member of the Advisory Committee that drafted the Federal Rules of Evidence during the early 1970s, and a critic of the partisan deployment of science in the adversarial courtroom, Weinstein embraced the stringent 95% significance threshold as a ready-made admissibility test that could validate the veracity of the statistical evidence used in court. Thus, while he referred to epidemiological studies as “the best (if not the sole) available evidence in mass exposure cases,” he nevertheless refused to accept them in evidence, unless they were statistically significant.

In the absence of statistical significance, the veterans’ lawyers based their proof of factual causation on animal studies, and supported it with occupational studies of industrial accidents involving dioxin that demonstrated the potential of dioxins to cause many of the ailments involved. But Weinstein discounted both types of evidence. The differences in species tested and in the high levels of exposure examined, he maintained, undermined the significance of these studies, and without the support of epidemiology they did not suffice to prove causation in tort. Still, like Jenkins before him, Weinstein was reluctant to allow the strict views on causation in tort prevent the veterans from recovering. Unable to satisfy the stringent standard of proof required in tort, Weinstein chose to question the

---

63 Ref 54.
applicability of this standard, and by implication, the applicability of the entire traditional tort system to the late-modern phenomenon of mass toxic tort litigation.\textsuperscript{65}

Epidemiologists use a simple relative index to measure risk in exposure cases. The index is defined by the ratio of the measured incidents of the disease in the exposed (numerator) and unexposed (denominator) groups tested. A risk ratio of one signifies that the incidence rate is the same among the exposed and non-exposed and thus indicates a lack of association between the suspected exposure and the alleged disease; a risk ratio greater than one suggests that the exposed are in higher risk of disease than the non-exposed; a risk ratio greater than two indicates that the exposed more than doubled their chance to contract the disease. From a population perspective, this means that more than half of the exposed owed their disease to the exposure. From the individual’s perspective, the epidemiologists suggested, one could interpret it to mean that the exposure was more likely than not responsible for his or her specific disease.\textsuperscript{66}

To remind you, under the traditional causation doctrine in tort, statistical correlations alone were insufficient, even if indicating that the probability of causation exceeds fifty percent (e.g. a risk ratio greater than 2). Some additional proof was required to shift the legal mind to one side or the other; preferably some direct testimony about the causal relationship between the defendant's conduct and the plaintiff's injury. Weinstein agreed with Jenkins that the chance for such evidence is very small in mass toxic tort cases, and that the consequence of retaining this requirement might allow defendants whom, “it is virtually certain, have injured thousands of people and caused billions of dollars in damage to be free of liability.”\textsuperscript{67} Jenkins, in his bellwether cases, offered a weaker interpretation of the preponderance standard that allowed a verdict in mass tort cases chiefly on statistical evidence. Weinstein seemed to side with Jenkins’ new interpretation but he argued that its successful adoption required further procedural adjustments.

\textsuperscript{65} In re: Agent Orange Product Liability Litigation, 597 F. Sup. 749 (1984).
\textsuperscript{67} In re: Agent Orange Product Liability Litigation, 597 F. Sup. 749 (1984).
Weinstein pointed out that the application of epidemiological evidence in a mass tort action on a case-by-case basis will not only be an administrative nightmare but will almost always result in either under or over-compensation. If the probability calculated is a hair less than 50%, each and all plaintiffs will lose and a clearly tortuous defendant could walk away. And if probability be a hair over 50%, each and all plaintiffs will win, including those not injured by the defendant. Shifting the burden of proof does not solve the problem. A defendant would still have to compensate all or no one, depending on which side of the 50% threshold the probability fell. This made no sense to Judge Weinstein. Given the unprecedented scale of mass tort and its financial stakes, he was worried of the potential implications of this problem, which could lead to the financial ruin of an entire industry or the deprivation of a large number of injured people from proper compensation.⁶⁸

Weinstein’s solution was as straightforward as it was radical: given the necessarily heavy reliance on statistical evidence in mass exposure cases, the time-honored tort practices of plaintiff-by-plaintiff and winner-takes-it-all will have to go. Mass tort cases should “try all plaintiffs’ claims together in a class action thereby arriving at a single, class-wide determination of the total harm to the community of plaintiffs. . . The defendant would then be liable to each exposed plaintiff for a pro rate share of that plaintiff’s injuries.” In short, if mass toxic torts are to allow verdicts based on statistical evidence, the courts need to match it with the equally aggregative and innovative mechanisms (at least in tort) of class action and proportional liability.⁶⁹

Weinstein was aware that his cutting of the Gordian knot of mass toxic tort ran against the legal grain and would most probably fail if the Agent Orange action would go to trial. He therefore pushed the parties to sign an out-of-court agreement he engineered. He cajoled the industry to put together a modest $180 millions fund and ordered its distribution among the 250,000 Vietnam veterans on the degree of disability alone,

---

⁶⁸ Id.
⁶⁹ Id.
regardless of cause. True to his analysis, Weinstein later summarily dismissed without a trial the individual claims of those veterans who chose to opt out of the agreement. Under the existing tort doctrines, he ruled, it was unfeasible to causally connect their individual ailments to Agent Orange exposure without solid epidemiological evidence.  

*Allen v. United States* and *Agent Orange* were key chapters in the adaptation of late 20th century American tort law to the challenges of mass toxic tort litigation. They put on display the inadequacies of the traditional tort doctrine of causation in dealing with mass toxic torts litigation and clarified many of the differences between the questions asked by law and the answers given by science. Jenkins and Weinstein, each was able to fashion a remedial process to compensate for the evidentiary complexities inherent in mass toxic tort litigation. Neither of their solutions seemed general enough, but both pointed to the central role epidemiological evidence came to play in the resolution of mass toxic tort cases. Jenkins’ solution depended on the unlikely availability of an authoritative body of epidemiological research, but his case by case approach was inapplicable en mass. Weinstein’s solution, fusing probabilistic causation with class action and proportional liability, was equally inapplicable for the everyday businesses of toxic torts. Nevertheless, his dismissal of animal studies as of “so little probative force and are so potentially misleading as to be inadmissible;” and his championing of epidemiology and its strict statistical-significance test proved remarkably influential in the years to come.  

**Junk science, epidemiology, and a legal reform**

The rapid growth of mass tort litigation and its financial consequences bred much anxiety and contention.  

Not surprisingly perhaps, the bulk of the criticism was directed at the science involved. By the early 1990s, the alarm was sounded that America’s courts were

---

being swamped by junk science, produced by unscrupulous experts hired by opportunistic attorneys aiming for the deep pockets of America’s corporations. The role of epidemiology was central to this growing debate over junk science. This time around, the critics were concerned less with the scientific nature of epidemiology and more with the ability of the courts to handle its ruse. Respectable judges found themselves more confused than enlightened by technical terms such as P-values, confidence margins, and significant levels, and made embarrassing mistakes. Still, judges could be trained and procedures could be improved. The real concern lay with the lay jury and their ability to handle the rich subtleties produced by the exploding market of expert epidemiological advice. The distrust in the jury’s capacity to handle complex evidence runs long and deep in American legal culture and the well-financed junk science campaign gave it new energy and focus. To shield the credulous jury from pseudoscientific expertise and protect corporate America from greedy lawyers, the judges were urged to become more vigilant with the new science they let into their court.

The complimentary debates about the proper role of judges and epidemiologists in mass tort litigation, and the standards each of them should follow in their own art have crossed paths in another mass toxic tort litigation that has occupied the courts since the early 1980s and involved Bendectin, a drug that was widely prescribed during the 1960s and 1970s for pregnant women to combat nausea. Ultimately, approximately 2,000 suits were filed against the drug manufacturer, Merrell-Dow Pharmaceuticals, Inc., asserting that

---


77 see refs 72 and 73
Bendectin caused a wide variety of birth defects, ranging from limb reductions to heart defects to neurological problems.\textsuperscript{78}

Like in other mass toxic tort litigations, the crucial battles of the Bendectin litigation were over the causal relation between Bendectin and the plaintiffs’ illnesses. To prove a causal link, the plaintiffs offered toxicological evidence that was based on in-\textit{vivo} and in \textit{vitro} studies that found links between Bendectin and malformation; and on chemical analysis that pointed to structural similarities between Bendectin and other substances known to cause birth defects. On the other side, Merrell-Dow’s lawyers based their defense strategy on the strength of a growing number of epidemiological studies that have failed to demonstrate a statistically-significant causal connection. Quoting Weinstein’s in \textit{Agent Orange}, they discounted the relevancy of the animal studies and chemical analysis, and claimed that in the absence of solid epidemiological support they were insufficient to show causation in tort.\textsuperscript{79}

By the end of the 1980s, the ongoing evidentiary battle between the plaintiffs and the Merrell-Dow in the Bendectin litigation was tilting towards the later.\textsuperscript{80} Finding themselves increasingly dependent on epidemiological evidence, the courts responded to the growing criticism against junk science by adopting the 95\% statistical-significance test as an admissibility threshold. This capped the remarkable change in the legal status of epidemiology. If at the start of the 1980s the courts still debated whether to allow epidemiologists to testify to the issue of causation, then by the start of the 1990s they


\textsuperscript{79} The first case was Mekdeci v. Merrell Nat'l Lab., 711 F.2d 1510 (11th Cir. 1983). The second plaintiffs’ verdict was Oxendine v. Merrell Dow Pharmaceuticals, 506 A.2d 1100 (D.C. 1986). A mass trial was conducted in Cincinnati in 1985 on behalf of over 1,000 infants, on the sole question of Bendectin's capacity to cause eight broad classes of birth defects. See In re Richardson-Merrell, Inc. "Bendectin" Prods. Liab. Litig., 624 F. Supp. 1212 (S.D. Ohio 1985).

were already summarily dismissing suits, and even reversing jury verdicts, because they
could not support themselves with statistically-significant epidemiological evidence.\footnote{81}

Epidemiology was not given a free hand in the courtroom. To fit it into tort, the courts
divided the proof of causation to two: general and specific. General causation referred the
potential of a given exposure to cause injury; specific causation, to the actual harm
claimed by the plaintiff. This partition of causality was rooted in the politics of the jury
system and the growing concern with the ability of the lay jury to handle the complexities
of the scientific evidence in late 20\textsuperscript{th}-century. The proof of general causation,
increasingly provided by epidemiology, is checked by the judge before the trial, during
the admissibility stage.\footnote{82} Only upon the judge’s satisfaction that the potential for harm
was proven, could the legal action move forward to the trial stage, where the issue of
specific causation can be examined by the jury.\footnote{83}

**Daubert and the legal standards of admissibility**

Prior to the twentieth century, there was no special admissibility test for scientific
evidence. Like every other type of evidence, it was mainly evaluated according to its
relevancy; helpfulness, and the qualifications of the witness.\footnote{84} Wary of the need to give
preference to one kind of science over another, 19\textsuperscript{th}-century judges followed a lenient
admissibility policy in the case of expert witnesses, and left it for the lawyers to expose
quackery during cross-examination, and for the jury to be the judge of the ensuing battles

\footnotesize
\begin{itemize}
  \item \footnote{81} Id.
  \item \footnote{82} See Daubert vs. Merrell Dow Pharmaceuticals, 509 U.S. 579 (1993); Margaret Berger, What
  \item \footnote{83} Dillingham, William O; Hagan, Patrick J; Salas, Rodrigo, Blueprint for General Causation
Analysis in Toxic Tort Litigation, FDCC Quarterly (Oct 1, 2003)
  \item \footnote{84} C. T. McCormick, *McCormick’s Handbook of the Law of Evidence* (St. Paul: West Publishing
Co., 1972), 2\textsuperscript{nd} ed., 489.
\end{itemize}
between the lawyers and the experts.\textsuperscript{85} No one, of course, trusted the jury to be able to do this job properly. Still, the courts considered it a fair price to pay for a free market of expertise that was considered the best protection from the abuse of political and executive powers.\textsuperscript{86}

It was only after World War II that American courts began to consistently apply a distinct standard for the admissibility of scientific evidence, and even then it was only in criminal cases.\textsuperscript{87} To that end, the courts resurrected a 1923 opinion of the Court of Appeal of the District of Columbia that rationalized the decision of the lower court to exclude a prominent expert in scientific lie-detection from testifying in a murder case to the veracity of his client’s alibi. “While courts will go a long way in admitting expert testimony deduced from a well recognized scientific principle or discovery,” the DC Court of Appeal prescribed, “the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs.”\textsuperscript{88} The lie-detector technology, the appellate court reasoned, did not receive such general acceptance and was therefore properly excluded.

Known as the ‘general acceptance’ standard, or simply as \textit{Frye} (after the defendant’s name), the courts increasingly used it during the 1960s and 70s to decide the admissibility of an array of technologies that was offered by the up-and-coming crime laboratories: voice prints, neutron activation analysis, gunshot residue tests, bite mark comparisons, scanning electron microscopic analysis, truth sera, and others.\textsuperscript{89} By the 1980s, \textit{Frye} was already well established as the general standard for the admissibility of novel scientific evidence in criminal trials. Still, Frye was not the only user manual in

\textsuperscript{86} Golan, Id.
\textsuperscript{88} Frye v. United States, \textit{Federal Reports} 293 (1923), 1013-1014.
town. In 1975 the *Federal Rules of Evidence* (FRE) were enacted and prescribed no special test to ensure the reliability of scientific evidence, new or old. Instead the FRE cast the widest net possible and provided that “If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of opinion or otherwise.”\(^9^0\) The FRE was generally interpreted as more liberal of the two standards, encouraging a more flexible judicial consideration of scientific evidence. On the other hand, since the FRE did not state an explicit intent to abandon *Frye*, some federal, and almost all state courts, remained committed to the general acceptance criterion as the prerequisite to the admissibility of scientific evidence, at least in criminal cases.\(^9^1\)

The tensions between judges and experts, between laboratory and statistical science, and between *Frye* and the FRE, all came to a head in 1993, in yet another Bendectin case in which a minor named Jason Daubert sued Merrell Dow for his birth defects.\(^9^2\) Daubert’s lawyers offered the court the usual toxicological mix of in-vitro, in vivo, and structural evidence that pointed to links between Bendectin and birth defects. In light of the growing legal reliance on epidemiological evidence, Daubert’s lawyers were careful to support their cause with a well-qualified expert, who re-analyzed previously published epidemiological data and was able to detect statistically-significant links between the drug and birth defects. Alas, the trial judge rejected the proposed re-analysis. Prepared especially for the trial, he reasoned, the re-analysis was not subjected to peer-review and thus could not be considered under *Frye* as generally accepted. Consequently, without epidemiological support, the federal trial judge considered the rest of the plaintiffs’ scientific evidence insufficient to prove causation and gave a summary judgment for the defendant, Merrell Dow.\(^9^3\)


\(^9^1\) Golan, 261


\(^9^3\) Id.
Daubert’s lawyers appealed all the way to the Supreme Court, arguing that the FRE supersedes Frye and that according to the FRE, it is for a jury, not a judge, to determine the persuasiveness of their scientific evidence. The Supreme Court, which until then refused to address the evidentiary issues raised by toxic tort litigation, agreed to review the case in order to clarify the proper admissibility standard of scientific evidence in mass tort litigation. Upon review, the Supreme Court agreed with the petitioners that Frye was superseded by the FRE, but rejected their let-it-all-in interpretation of the FRE. Instead, the Supreme Court read the FRE as prescribing a new function for the trial judge – to ensure that the scientific evidence admitted into his courtroom is reliable. Addressing the main question left open by the FRE—how one recognizes reliable scientific knowledge—the Supreme Court equipped the trial judge with a classical flexible recipe of five factors that could be used in determining the quality of the scientific evidence proposed.94

1. Falsifiability and Testability: whether the theory or technique can be falsified and had been tested.
2. Peer Review: whether the theory or technique had been subjected to peer review.
3. Error rates: known or potential error rate
5. General Acceptance (the Frye test): the degree to which the theory or technique has been accepted by the relevant scientific community.

The Supreme Court’s list of five factors, none of which is sufficient or necessary, resembles in structure and purpose the shopping lists epidemiologists developed to decide causation. - the most famous of which is probably the 1965 nine-factors list suggested by Sir Bradford Hill. 95 But perhaps we should not be surprised by this resemblance since, as I

95 Bradford Hill, The environment and disease: association or causation? Proceedings of the Royal Society of Medicine (1965) 58:295–300. The list included strength, consistency and
tried to demonstrate in this paper, the epidemiological quest for true causes and the legal quest for true science have developed hand-in-hand, as part of the evolving interplay of theory, practices and politics in the late-modern adversarial courtroom, where true causes are hard to find, no single explanatory factor is enough, and neither the partisan expert nor the lay juror can be trusted to decide the evidence alone.

specificity of the associations; temporal and dose-response relations; biological plausibility, experimental support, helpful analogies, and general coherence..