PAPER TITLE: The Convergence of Law and Science on Epistemic Virtues and Vices*

Nathan A. Schachtman
nschachtmn@ubglaw.com

SESSION TITLE: Fraud, Error, and Ethics Violations: Exposing Bad Science in Litigation

Paper Presented by:
Nathan A. Schachtman, Of Counsel, UB Greensfelder

For over 40 years, Nathan has been trying to make American court rooms safe for manufacturers. He has counseled clients on strategies to avoid, resolve, and win litigation. He has served as national coordinating, science, MDL, trial, and appellate counsel in product liability, pharmaceutical, occupational disease, and environmental cases. He has tried dozens of cases in New York, New Jersey, and Pennsylvania — where he has been admitted — as well as in other states, *pro hac vice*. He has been involved in many of the leading American cases involving epidemiologic and other scientific evidence in product liability cases.
.

Nathan is an elected member of the American Law Institute, and a Life Fellow of the American Bar Foundation. Has has taught statistics and probability at the Columbia Law School, as a lecturer in law. He currently serves on the Board of Directors for the Center for Truth in Science.

*
Disclaimer: This paper benefited from the comments and corrections of my co-panelists, Drs. Erica James and Ivan Oransky. The errors that remain are solely my own. The author, as defense counsel, was a participant for some of the proceedings described in this paper. The views and opinions presented in this paper are those of the author, and not necessarily shared by his clients, legal colleagues, friends, and family, although they should be. Facts on the other hand are stubborn things, and we are all stuck with them.

Table of Contents

I. Introduction

The law of expert witness opinion underwent an abrupt shift, in 1975, with the adoption of an epistemic standard for admissibility in Rule 702, of the Federal Rules of Evidence. Before Rule 702, the common law generally required only that witnesses be qualified, and that their opinion be relevant. In 1923, one federal appellate court adopted the "twilight zone" test of general acceptance,[1] which was essentially a sociological test. Many, if not most, courts restricted this general acceptance test to opinions based upon novel devices, such as lie detectors. After the adoption of Rule 702, the federal courts were slow to recognize the full implications of its epistemic test for admissibility. Retrograde decisions based upon mere qualifications and relevancy abounded, but the last 30 years have seen the slow but steady elimination easy admissibility. The "liberal thrust" of the federal rules has largely been achieved by pushing out authoritarian opinion propped up by the false allure of credentials, and lacking sufficient support with valid evidence and inference.

The health sciences, which invariably provide the scientific research that is the basis for health claims in court, have themselves undergone a remarkable transformation in the last century. Public health in the early 20th century was largely focused on infectious disease, and the search for a necessary (but not always sufficient) pathogen. After World War II, and the introduction of antibiotics and vaccinations, public health dramatically shifted its focus to chronic diseases, which often involved causes that could be inferred only by epidemiologic methods, with the assistance of sound statistical analysis.[2] The introduction of statistical epidemiologic methods had notably successes, especially with the introduction of methods to address confounding.[3]

The sophistication of the health sciences increased in the 1970s, with the development of systematic reviews, meta-analyses, refined statistical techniques, and improved approaches for identifying systematic biases and confounding. There were retrograde movements in the health sciences as well, recognized through studies that could not be reproduced, and the identification of prevalent questionable research practices that undermined the validity of study results and causal inferences. Retraction of study articles, a rarity in the 20th century, has become a commonplace in this century.[4]

---

[1] *Frye v. United States*, 293 F. 1013, 1014 (D.C. Cir. 1923).

[2] Mark Parascandola, "The epidemiologic transition and changing concepts of causation and causal inference," 64 *Revue d'histoire des sciences* 243 (2011); Colin Talley, Howard I. Kushner & Claire E. Sterk, "Lung Cancer, Chronic Disease Epidemiology, and Medicine, 1948-1964," 59 *J. History Med. & Allied Sciences* 329 (2004).

[3] *See, e.g.*, Nathan Mantel & William Haenszel, "Statistical aspects of the analysis of data from retrospective studies of disease," 22 *J. Nat'l Cancer Instit.* 19 (1959). *See also* Mervyn Susser, "Epidemiology in the United States after World War II: The Evolution of Technique," 7 *Epidem. Rev.* 147 (1985).

[4] *See, e.g.*, Murat Cokol, Fatih Ozbay, and Raul Rodriguez-Esteban, "Retraction rates are on the rise," 9 *European Molecular Biol. Reports* 2 (2008).

The published article is the typical building block for expert witness opinion in litigation of health claims. Now that the admissibility standards for expert witness opinion testimony requires attention to the sufficiency of the expert witnesses' facts and data, to the validity of their methodology, and to the validity of their application of their methodology to the facts of cases, lawyers must pay attention to the presence of questionable research practices and become more sophisticated consumers of, and advocates for, good science. Law and science have converged in their concerns for validity concerns in studies, and in drawing causal conclusions.

II. *Nullius in verba*

The 1975 codification of the law of evidence, in the Federal Rules of Evidence, introduced a subtle, aspirational criterion for expert witness opinion – ***knowledge***. As originally enacted, Rule 702 read:

"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 an opinion or otherwise."[5]

In case anyone missed the point, the Advisory Committee Note for the original Rule 702 emphasized that they intended for the standard to be an epistemic standard:

> "An intelligent evaluation of facts is often difficult or impossible without the application of some scientific, technical, or other specialized ***knowledge***. The most common source of this ***knowledge*** is the expert witness, although there are other techniques for supplying it."[6]

Perhaps we should not be too surprised that the epistemic standard was missed by most judges, and even by most lawyers. For a very long time, the common law set out a minimal test for expert witness opinion testimony. The expert witness had to be qualified by training, experience, or education, and the opinion proffered had to be logically and legally relevant to the issues in the case.[7] The enactment of Rule 702, in 1975, barely made a dent in the regime of easy admissibility.

Before the Federal Rules of Evidence, there was, of course, the famous *Frye* case, which involved an appeal from the excluded expert witness opinion based upon William Marston's lasso of truth, the polygraph machine. In 1923, the court in *Frye* affirmed the exclusion of the expert witness opinion, based upon the lack of general acceptance of the device's reliability, with its famous twilight zone language:[8]

---

[5] Pub. L. 93–595, §1, Jan. 2, 1975, 88 Stat. 1937 (emphasis added).
[6] Notes of Advisory Committee on Proposed Rules (1975) (emphasis added).
[7] *See* Charles T. McCormick, *Handbook of the Law of Evidence* 28-29, 363 (1954) ("Any relevant conclusions which are supported by a qualified expert witness should be received unless there are other reasons for exclusion.")
[8] *Frye v. United States*, 293 F. 1013, 1014 (D.C. Cir. 1923).

"Just when a scientific principle or discovery crosses the line between the experimental and demonstrable stages is difficult to define. Somewhere in this twilight zone the evidential force of the principle must be recognized, and while courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, 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."

With the explosion of tort litigation fueled by strict products liability doctrine, lawyers pressed *Frye*'s requirement of general acceptance into service as a bulwark against unreliable scientific opinions. Many courts, however, limited *Frye* to novel devices, and in 1993, the Supreme Court, in *Daubert*,[9] rejected the legal claim that Rule 702 had incorporated the common law "general acceptance" test. Looking to the language of the rule itself, the Supreme Court discerned that the rule laid down an epistemic test, not a call for sociological surveys about the prevalence of beliefs.

Resistance to the spirit and text of Rule 702 has been widespread and deep seated. Several years after the adoption of Rule 702, the Court of Appeals, in a chemical exposure case, expressed a standard that encouraged a willingness to disregard epistemic requirements in favor of naked expertise and bare relevance:

> "On questions such as these, which stand at the frontiers of current medical and epidemiological inquiry, if experts are willing to testify that such a link exists, it is for the jury to decide whether to credit such testimony."[10]

The *Ferebee* articulation of common law laissez faire was largely banished from federal court by the *Daubert* decision. After *Daubert*, the Supreme Court decided three more cases to emphasize that the epistemic standard was "exacting" and that it would not go away.[11] In 2000, in the wake of the Supreme Court's quartet of decisions, Rule 702 was amended substantively to incorporate some of the essence of the Supreme Court's observations about the necessary requirements for the admissibility of expert witness opinion testimony,[12] such as the requirement that the

---

[9] *Daubert v. Merrell Dow Pharm., Inc*., 509 U.S. 579 (1993).

[10] *Ferebee v. Chevron Co*., 736 F.2d 1529, 1534 (D.C. Cir.) (affirming the rejection of defendant's *Frye* challenge), *cert. denied*, 469 U.S. 1062 (1984). I have argued elsewhere that the scientific basis for Ferebee's claim may well have been better than suggested in the quote above. Schachtman, "Ferebee Revisited," *Tortini* (Dec. 28, 2017).

[11] *General Electric Co. v. Joiner*, 522 U.S. 136 (1997); *Kumho Tire Co. v. Carmichael*, 526 U.S. 137 (1999); *Weisgram v. Marley Co*., 528 U.S. 440 (2000). In *Kumho Tire*, Justice Breyer, writing for the court, gave an alternative expression to the clearly epistemic goals of Rule 702, which "is to make certain that an expert, whether basing testimony upon professional studies or personal experience, employs in the courtroom the same level of intellectual rigor that characterizes the practice of an expert in the relevant field." 526 U.S. at 152. This articulation ensures that the standards of the discipline are imported into the admissibility determinations by courts.

[12] *See* notes 5, 6, *supra*.

proponent of the opinion establish that it is based upon sufficient facts or data, is the product of reliable principles and methods, and is the result of reliably applying those reliable principles and methods to the facts of the case.

The change in the law of expert witnesses, in the 1990s, left some academic commentators well-nigh apoplectic. One professor of evidence law at a large law school complained that the law was a "conceptual muddle containing within it a threat to liberty and popular participation in government."[13] Many federal district and intermediate appellate courts responded by ignoring the language of Rule 702, by reverting to pre-*Daubert* precedent, or by inventing new standards and shifting the burden to the party challenging the expert witness opinion's admissibility. For many commentators, lawyers, and judges, science had no validity concerns that the law was bound to respect.

The judicial evasion and avoidance of the requirements of Rule 702 did not go unnoticed. Professor David Bernstein and practicing lawyer Eric Lasker wrote a paper in 2015, to call attention to the judicial disregard of the requirements of Rule 702.[14] In the wake of this paper, several years of discussion and debate ensued before the Judicial Conference Advisory Committee on Evidence Rules (AdCom). In 2021, the AdCom documented that "in a fair number of cases, the courts have found expert testimony admissible even though the proponent has not satisfied the Rule 702(b) and (d) requirements by a preponderance of the evidence."[15] This frank acknowledgment led the AdCom to propose amending Rule 702, "to clarify and emphasize" that gatekeeping requires determining whether the proponent has demonstrated to the court "that it is more likely than not that the proffered testimony meets the admissibility requirements set forth in the rule."[16] The Proposed Committee Note written in support of amending Rule 702 observed that "many courts have held that the critical questions of the sufficiency of an expert's basis, and

---

[13] John H. Mansfield, "An Embarrassing Episode in the History of the Law of Evidence," 34 *Seton Hall L. Rev*. 77, 77 (2003); *see also* John H. Mansfield, "Scientific Evidence Under *Daubert*," 28 *St. Mary's L.J*. 1, 23 (1996). Professor Mansfield was the John H. Watson, Jr., Professor of Law, at the Harvard Law School. Many epithets were thrown in the heat of battle to establish meaningful controls over expert witness testimony. *See, e.g*., Kenneth Chesebro, "Galileo's Retort: Peter Huber's Junk Scholarship," 42 *Am. Univ. L. Rev*. 1637 (1993). Mr. Chesebro was counsel of record for plaintiffs-appellants in *Daubert*, well before he became a convicted racketeer in Georgia.

[14] David Bernstein & Eric Lasker, "Defending Daubert: It's Time to Amend Federal Rules of Evidence 702," 57 *Wm. & Mary L Rev*. 1 (2015).

[15] Report of AdCom (May 15, 2021), at https://www.uscourts.gov/rules-policies/archives/committee-reports/advisory-committee-evidence-rules-may-2021. *See also* AdCom, Minutes of Meeting at 4 (Nov. 13, 2020) ("[F]ederal cases . . . revealed a pervasive problem with courts discussing expert admissibility requirements as matters of weight.")], at https://www.uscourts.gov/rules-policies/archives/meeting-minutes/advisory-committee-evidence-rules-november-2020.

[16] Proposed Committee Note, Summary of Proposed New and Amended Federal Rules of Procedure (Oct. 19, 2022), at https://www.uscourts.gov/sites/default/files/2022_scotus_package_0.pdf

the application of the expert's methodology, are questions of weight and not admissibility. *These rulings are an incorrect application of Rules 702 and 104(a).*"[17]

The proposed new Rule 702 is now law,[18] with its remedial clarification that the proponent of expert witness opinion must show the court that the opinion is sufficiently supported by facts or data,[19] that the opinion is "the product of reliable principles and methods,"[20] and that the opinion "reflects a reliable application of the principles and methods to the facts of the case."[21] The Rule prohibits deferring the evaluation of sufficiency of support or reliability of application of method to the trier of fact; there is no statutory support for suggesting that these inquires always or usually go to "weight and not admissibility," or that there is a presumption of admissibility.

We may not have reached the Age of Aquarius, but the days of "easy admissibility" should be confined to the dustbin of legal history. Rule 702 is quickly approaching its 50th birthday, with the last 30 years witnessing the implementation of the promise and potential of an epistemic standard of trustworthiness for expert witness opinion testimony. Rule 702, in its present form, should go a long way towards putting validity questions squarely before the court under Rule 702. *Nullius in verba*[22] has been the motto of the Royal Society since 1660; it should now guide expert witness practice in federal court going forward.

III. QRPs in Science and in Court

Lay juries usually function well in assessing the relevance of an expert witness's credentials, experience, command of the facts, likeability, physical demeanor, confidence, and ability to communicate. Lay juries can understand and respond to arguments about personal bias, which no doubt is why trial lawyers spend so much time and effort to emphasize the size of fees and consulting income, and the propensity to testify only for one side. For procedural and practical reasons, however, lay juries do not function very well in assessing the actual merits of scientific controversies. And with respect to methodological issues that underlie the merits, juries barely function at all. The legal system imposes no educational or experiential qualifications for jurors, and trials are hardly the occasion to teach jurors the methodology, skills, and information needed to resolve methodological issues that underlie a scientific dispute.

---

[17] *Id*. (emphasis added).
[18] In April 2023, Chief Justice Roberts transmitted the proposed Rule 702, to Congress, under the Rules Enabling Act, and highlighted that the amendment "shall take effect on December 1, 2023, and shall govern in all proceedings thereafter commenced and, insofar as just and practicable all proceedings then pending." S. Ct. Order, at 3 (Apr. 24, 2023), https://www.supremecourt.gov/orders/courtorders/frev23_5468.pdf; S.Ct. Transmittal Package (Apr. 24, 2023), < https://www.uscourts.gov/sites/default/files/2022_scotus_package_0.pdf>.
[19] Rule 702(b).
[20] Rule 702(c).
[21] Rule 702(d).
[22] Take no one's word for it.

Scientific studies, reviews, and meta-analyses are virtually never directly admissible in evidence in courtrooms in the United States. As a result, juries do not have the opportunity to read and ponder the merits of these sources, and assess their strengths and weaknesses. The working assumption of our courts is that juries are not qualified to engage directly with the primary sources of scientific evidence, and so expert witnesses are called upon to deliver opinions based upon a scientific record not directly in evidence. In the litigation of scientific disputes, our courts thus rely upon the testimony of so-called expert witnesses in the form of opinions. Not only must juries, the usual trier of fact in our courts, assess the credibility of expert witnesses, but they must assess whether expert witnesses are accurately describing studies that they cannot read in their entirety.

The convoluted path by which science enters the courtroom supports the liberal and robust gatekeeping process outlined under Rules 702 and 703 of the Federal Rules of Evidence. The court, not the jury, must make a preliminary determination, under Rule 104, that the facts and data of a study are reasonably relied upon by an expert witness (Rule 703). And the court, not the jury, again under Rule 104, must determine that expert witnesses possess appropriate qualifications for relevant expertise, and that these witnesses have proffered opinions sufficiently supported by facts or data, based upon reliable principles and methods, and reliably applied to the facts of the case. (Rule 702). There is no constitutional right to bamboozle juries with inconclusive, biased, confounded, or crummy studies, or selective and incomplete assessments of the available facts and data. Back in the days of "easy admissibility," opinions could be tested on cross-examination, but limited time and acumen of counsel, court, and juries cry out for meaningful scientific due process along the lines set out in Rules 702 and 703.

The evolutionary development of Rules 702 and 703 has promoted a salutary convergence between science and law. According to one historical overview of systematic reviews in science, the foundational period for such reviews (1970-1989) overlaps with the enactment of Rules 702 and 703, and the institutionalization of such reviews (1990-2000) coincides with the development of these Rules in a way that introduced some methodological rigor into scientific opinions that are admitted into evidence.[23]

The convergence between legal admissibility and scientific validity considerations has had the further result that scientific concerns about the quality and sufficiency of underlying data, about the validity of study design, analysis, reporting, and interpretation, and about the adequacy and validity of data synthesis, interpretation, and conclusions have become integral to the gatekeeping process. This convergence has the welcome potential to keep legal judgments more in line with best scientific evidence and practice.

---

[23] Quan Nha Hong & Pierre Pluye, "Systematic Reviews: A Brief Historical Overview," 34 *Education for Information* 261 (2018); Mike Clarke & Iain Chalmers, "Reflections on the history of systematic reviews," 23 *BMJ Evidence-Based Medicine* 122 (2018); Cynthia Farquhar & Jane Marjoribanks, "A short history of systematic reviews," 126 *Brit. J. Obstetrics & Gynaecology* 961 (2019); Edward Purssell & Niall McCrae, "A Brief History of the Systematic Review," chap. 2, in Edward Purssell & Niall McCrae, *How to Perform a Systematic Literature Review: A Guide for Healthcare Researchers, Practitioners and Students* 5 (2020).

The science-law convergence also means that courts must be apprised of, and take seriously, the problems of study reproducibility, and more broadly, the problems raised by questionable research practices (QRPs), or what might be called the patho-epistemology of science. The development, in the 1970s, and the subsequent evolution of the systematic review represented the scientific community's rejection of the old-school narrative reviews that selected a few of all studies to support a pre-existing conclusion. Similarly, the scientific community's embarrassment, in the 1980s and 1990s, over the irreproducibility of study results, has in this century grown into an existential crisis over study reproducibility in the biomedical sciences.

In 2005, John Ioannidis published an article that brought the concern over "reproducibility" of scientific findings in bio-medicine to an ebullient boil.[24] Ioannidis pointed to several factors, which alone or in combination, in his view, rendered most published medical findings likely false. Among the publication practices responsible for this unacceptably high error rate, Ioannidis identified the use of small sample sizes, data-dredging and p-hacking techniques, poor or inadequate statistical analysis, undue flexibility in research design, conflicts of interest, motivated reasoning, fads, and prejudices, and pressure to publish "positive" results. The results, often with small putative effect sizes, across an inadequate number of studies, are then hyped by lay and technical media, as well as the public relations offices of universities and advocacy groups, only to be further misused by advocates, and further distorted to serve the goals of policy wonks. Social media then reduces all the nuances of a scientific study to an insipid meme.

Ioannidis' critique resonated with lawyers. Legal practitioners in health effects litigation are no strangers to dubious research methods, lack of accountability, herd-like behavior, and a culture of generating positive results, often motivated by political or economic sympathies. Although lawyers must prepare for confronting dodgy methods in front of jury, asking for scientific due process that intervenes and decides the methodological issues with well-reasoned, written opinions in advance of trial does not seem like asking very much.

The sense that we are awash in false-positive studies was heightened by subsequent papers. In 2011, Uri Simonsohn and others showed that by using simulations of various combinations of QRPs in psychological science, researchers could attain a 61% false-positive rate for research outcomes.[25] The following year saw scientists at Amgen attempt replication of 53 important studies in hematology and oncology. They succeeded in replicated only six.[26] Also in 2012, Dr. Janet Woodcock, director of the Center for Drug Evaluation and Research at the Food and Drug Administration, "estimated that as much as 75 per cent of published biomarker associations are

[24] John P. A. Ioannidis "Why Most Published Research Findings Are False," 1 *PLoS Med* 8 (2005).
[25] Joseph P. Simmons, Leif D. Nelson, and Uri Simonsohn, "*False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant*," 22 *Psychological Sci.* 1359 (2011).
[26] C. Glenn Begley and Lee M. Ellis, "Drug development: Raise standards for preclinical cancer research," 483 *Nature* 531 (2012).

not replicable."[27] In 2016, the journal *Nature* reported that over 70% of scientists who responded to a survey had unsuccessfully attempted to replicate another scientist's experiments, and more than half failed to replicate their own work.[28] Of the respondents, 90% agreed that there was a replication problem. A majority of the 90% believed that the problem was significant.

The scientific community reacted to the perceived replication crisis in a variety of ways, from conceptual clarification of the very notion of reproducibility,[29] to identification of improper uses and interpretations of key statistical concepts,[30] to guidelines for improved conduct and reporting of studies.[31]

---

[27] Edward R. Dougherty, "Biomarker Development: Prudence, risk, and reproducibility," 34 *Bioessays* 277, 279 (2012); Turna Ray, "FDA's Woodcock says personalized drug development entering 'long slog' phase," *Pharmacogenomics Reporter* (Oct. 26, 2011).
[28] Monya Baker, "Is there a reproducibility crisis," 533 *Nature* 452 (2016).
[29] Steven N. Goodman, Daniele Fanelli, and John P. A. Ioannidis, "What does research reproducibility mean?," 8 *Science Translational Medicine* 341 (2016); Felipe Romero, "Philosophy of science and the replicability crisis," 14 *Philosophy Compass* e12633 (2019); Fiona Fidler & John Wilcox, "Reproducibility of Scientific Results," *Stanford Encyclopedia of Philosophy* (2018), available at https://plato.stanford.edu/entries/scientific-reproducibility/.
[30] Andrew Gelman and Eric Loken, "The Statistical Crisis in Science," 102 *Am. Scientist* 460 (2014); Ronald L. Wasserstein & Nicole A. Lazar, "The ASA's Statement on p-Values: Context, Process, and Purpose," 70 *The Am. Statistician* 129 (2016); Yoav Benjamini, Richard D. DeVeaux, Bradly Efron, Scott Evans, Mark Glickman, Barry Braubard, Xuming He, Xiao Li Meng, Nancy Reid, Stephen M. Stigler, Stephen B. Vardeman, Christopher K. Wikle, Tommy Wright, Linda J. Young, and Karen Kafadar, "The ASA President's Task Force Statement on Statistical Significance and Replicability," 15 *Annals of Applied Statistics* 1084 (2021).
[31] The International Society for Pharmacoepidemiology issued its first Guidelines for Good Pharmacoepidemiology Practices in 1996. The most recent revision, the third, was issued in June 2015. *See* "The ISPE Guidelines for Good Pharmacoepidemiology Practices (GPP)," available at https://www.pharmacoepi.org/resources/policies/guidelines-08027/. *See also*
"The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement Guidelines for Reporting Observational Studies
Erik von Elm, Douglas G. Altman, Matthias Egger, Stuart J. Pocock, Peter C. Gøtzsche, and Jan P. Vandenbroucke, for the STROBE Initiative," 18 *Epidem*. 800 (2007); Jan P. Vandenbroucke, Erik von Elm, Douglas G. Altman, Peter C. Gøtzsche, Cynthia D. Mulrow, Stuart J. Pocock, Charles Poole, James J. Schlesselman, and Matthias Egger, for the STROBE initiative, "Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration," 147 *Ann. Intern. Med*. W-163 (2007); Shah Ebrahim & Mike Clarke, "STROBE: new standards for reporting observational epidemiology, a chance to improve," 36 *Internat'l J. Epidem*. 946 (2007); Matthias Egger, Douglas G. Altman, and Jan P Vandenbroucke of the STROBE group, "Commentary: Strengthening the reporting of observational epidemiology—the STROBE statement," 36 *Internat'l J. Epidem*. 948 (2007).

Entire books dedicated to identifying the sources of, and the correctives for, undue researcher flexibility in the design, conduct, and analysis of studies, have been published.[32] In some ways, the Rule 702 and 703 case law is like the collected works of the Berenstain Bears, on how not to do studies. The consequences of the replication crisis are real and serious. Badly conducted and interpreted science leads to research wastage,[33] loss of confidence in scientific expertise,[34] contemptible legal judgments, and distortion of public policy.

The proposed correctives to QRPs deserve the careful study of lawyers and judges who have a role in health effects litigation.[35] Whether as the proponent of an expert witness, or the challenger, several of the recurrent proposals, such as the call for greater data sharing and pre-registration of protocols and statistical analysis plans,[36] have real-world litigation salience. In many instances, they can and should direct lawyers' efforts at discovery and challenging of the relied upon scientific studies in litigation.

IV. Peer Review, Protocols, and QRPs

A. Peer Review

In *Daubert*, the Supreme Court decided a legal question about the proper interpretation of a statute, Rule 702, and then remanded the case to the Ninth Circuit of the Court of Appeals for

---

[32] *See, e.g.*, Lee J. Jussim, Jon A. Krosnick, and Sean T. Stevens, eds., *Research Integrity: Best Practices for the Social and Behavioral Sciences* (2022); Joel Faintuch & Salomão Faintuch, eds., *Integrity of Scientific Research: Fraud, Misconduct and Fake News in the Academic, Medical and Social Environment* (2022); William O'Donohue, Akihiko Masuda & Scott Lilienfeld, eds., *Avoiding Questionable Research Practices in Applied Psychology* (2022); Klaas Sijtsma, *Never Waste a Good Crisis: Lessons Learned from Data Fraud and Questionable Research* Practices (2023).

[33] *See, e.g.*, Iain Chalmers, Michael B Bracken, Ben Djulbegovic, Silvio Garattini, Jonathan Grant, A Metin Gülmezoglu, David W Howells, John P A Ioannidis, and Sandy Oliver, "How to increase value and reduce waste when research priorities are set," 383 *Lancet* 156 (2014); John P A Ioannidis, Sander Greenland, Mark A Hlatky, Muin J Khoury, Malcolm R Macleod, David Moher, Kenneth F Schulz, and Robert Tibshirani, "Increasing value and reducing waste in research design, conduct, and analysis," 383 *Lancet* 166 (2014).

[34] *See, e.g.*, Friederike Hendriks, Dorothe Kienhues, and Rainer Bromme, "Replication crisis = trust crisis? The effect of successful vs failed replications on laypeople's trust in researchers and research," 29 *Public Understanding Sci*. 270 (2020).

[35] R. Barker Bausell, *The Problem with Science: The Reproducibility Crisis and What to Do About It* (2021).

[36] *See, e.g.*, Brian A. Noseka, Charles R. Ebersole, Alexander C. DeHavena, and David T. Mellora, "The preregistration revolution," 115 *Proc. Nat'l Acad. Soc.* 2600 (2018); Michael B. Bracken, "Preregistration of Epidemiology Protocols: A Commentary in Support," 22 *Epidemiology* 135 (2011); Timothy L. Lash & Jan P. Vandenbroucke, "Should Preregistration of Epidemiologic Study Protocols Become Compulsory? Reflections and a Counterproposal," 23 *Epidemiology* 184 (2012).

further proceedings. The Court did, however, weigh in with dicta about some several considerations in admissibility decisions. In particular, the Court identified four non-dispositive factors: whether the challenged opinion has been empirically tested, whether it had been published and peer reviewed, and whether the underlying scientific technique or method supporting the opinion has an acceptable rate of error, and whether it has gained general acceptance.[37]

The context in which peer review was discussed in *Daubert* is of some importance to understanding why the Court held out peer review as a consideration. One of the bases for the defense challenges to some of the plaintiffs' expert witnesses' opinions in *Daubert* was their reliance upon re-analyses of published studies to suggest that there was indeed an increased risk of birth defects if only the publication authors had used some other control group, or taken some other analytical approach. Re-analyses can be important, but these reanalyses of published Bendectin studies were *post hoc*, litigation driven, and obviously result oriented. The Court's discussion of peer review reveals that it was not simply creating a box to be checked before a trial court could admit an expert witness's opinions. Peer review was suggested as a consideration because:

> "submission to the scrutiny of the scientific community is a component of 'good science', in part because it increases the likelihood that ***substantive flaws in methodology will be detected***. The fact of publication (or lack thereof) in a peer reviewed journal thus will be a relevant, though not dispositive, consideration in ***assessing the scientific validity of a particular technique or methodology on which an opinion is premised***."[38]

Peer review, or the lack thereof, for the challenged expert witnesses' re-analyses was called out because it raised suspicions of lack of validity. Nothing in *Daubert*, in later decisions, or more importantly in Rule 702 itself, supports admitting expert witness testimony just because the witness relied upon peer-reviewed studies, especially when the studies are invalid or are based upon questionable research practices. The Court was careful to point out that peer-reviewed publication was "not a *sine qua non* of admissibility; it does not necessarily correlate with reliability, … ."[39] The Court thus showed that it was well aware that well-grounded (and thus admissible) opinions may not have been previously published, and that the existence of peer review was simply a potential aid in answering the essential question, whether the proponent of a proffered opinion has shown "the scientific validity of a particular technique or methodology on which an opinion is premised."[40]

Since 1993, much has changed in the world of bio-science publishing. The wild proliferation of journals, including predatory and "pay-to-play" journals, has disabused most observers that peer review provides evidence of validity of methods. Along with the exponential growth in publications has come an exponential growth in expressions of concern and out-right retractions

---

[37] *Daubert v. Merrell Dow Pharms., Inc*., 509 U.S. 579, 593-594 (1993).
[38] *Id*. at 594 (internal citations omitted) (emphasis added).
[39] *Id.*
[40] *Id*. at 593-94.

of articles, as chronicled and detailed at *Retraction Watch*.[41] Some journals encourage authors to nominate the peer reviewers for their manuscripts; some journals let authors block some scientists as peer reviewers of their submitted manuscripts. If the Supreme Court were writing today, it might well note that peer review is often a feature of bad science, advanced by scientists who know that peer-reviewed publication is the price of admission to the advocacy arena.[42]

Since the Supreme Court decided *Daubert*, the Federal Judicial Center and National Academies of Science have provided a *Reference Manual for Scientific Evidence*, now in its third edition, and with a fourth edition on the horizon, to assist judges and lawyers involved in the litigation of scientific issues. Professor Goodstein, in his chapter "How Science Works," in the third edition, provides the most extensive discussion of peer review in the *Manual*, and emphasizes that peer review "works very poorly in catching cheating or fraud."[43] Goodstein invokes his own experience as a peer reviewer to note that "peer review referees and editors limit their assessment of submitted articles to such matters as style, plausibility, and defensibility; they do not duplicate experiments from scratch or plow through reams of computer-generated data in order to guarantee accuracy or veracity or certainty."[44] Indeed, Goodstein's essay in the *Reference Manual* characterizes the ability of peer review to warrant study validity as a "myth":

> "***Myth***: The institution of peer review assures that all published papers are sound and dependable.
> ***Fact***: Peer review generally will catch something that is completely out of step with majority thinking at the time, but it is *practically useless for catching outright fraud*, and it is *not very good at dealing with truly novel ideas*. …
> It ***certainly does not ensure that the work has been fully vetted in terms of the data analysis and the proper application of research methods***."[45]

Goodstein's experience as a peer reviewer is hardly idiosyncratic. One standard text on the ethical conduct of research reports that peer review is often ineffective or incompetent, and that it may not even catch simple statistical or methodological errors.[46] According to the authors, Shamoo and Resnik:

> "[p]eer review is not good at detecting data fabrication or falsification partly because reviewers usually do not have access to the material they would need to detect fraud, such as the original data, protocols, and standard operating procedures."[47]

---

[41] *Retraction Watch*, at https://retractionwatch.com/.
[42] Drummond Rennie, "Guarding the guardians: a conference on editorial peer review," 256 *J. Am. Med. Ass'n* 2391, 2391 (1986).
[43] *Reference Manual on Scientific Evidence* at 37, 44-45 (3rd ed. 2011) [*Manual*].
[44] *Id*. at 44-45 n.11.
[45] *Id*. at 48 (emphasis added).
[46] Adil E. Shamoo and David B. Resnik, *Responsible Conduct of Research* 133 (4th ed. 2022).
[47] *Id*.

In 2008, the editors of the *British Medical Journal* put the effectiveness of statistical peer review to an empirical test by sending out papers seeded with major and minor statistical errors. On average, the 600 reviewers found three or fewer of the nine major errors. The editors concluded that they must not assume that reviewers will find most major errors.[48]

Without access to protocols, statistical analysis plans, and original data, peer review often cannot identify good faith or negligent deviations from the standard of scientific care. There is some evidence to support this negative assessment of peer review from testing of the counter-factual. Reviewers were able to detect questionable, selective reporting when they had access to the study authors' research protocols.[49]

B. Study Protocol

The study protocol provides the scientific rationale for a study, clearly defines the research question and the data collection process, defines the key exposure and outcomes, and describes the methods to be applied, before commencing data collection.[50] The protocol also typically pre-specifies the statistical data analysis. The epidemiology chapter of the current edition of the *Reference Manual for Scientific Evidence* offers blandly only that epidemiologists attempt to minimize bias in observational studies with "data collection protocols."[51] Epidemiologists and statisticians are much clearer in emphasizing the importance, indeed the necessity, of having a study protocol before commencing data collection. Back in 1988, John Bailar and Frederick Mosteller explained that it was critical in reporting statistical analyses to inform readers about how and when the authors devised the study design, and whether they set the design criteria out in writing before they began to collect data.[52]

---

[48] Sara Schroter, Nick Black, Stephen Evans, Fiona Godlee, Lyda Osorio, and Richard Smith, "What errors do peer reviewers detect, and does training improve their ability to detect them?" 101 *J. Royal Soc'y Med*. 507 (2008). *See also* Douglas Altman, "Statistics in medical journals: developments in the 1980s," 10 *Stat. Med*. 1897 (1991); Stuart J. Pocock, Michael D. Hughes, and Robert J. Lee, "Statistical problems in the reporting of clinical trials. A survey of three medical journals," 317 *New Engl. J. Med*. 426 (1987); Sheila M. Gore, Ian G. Jones, Eilef C. Rytter, "Misuse of statistical methods: critical assessment of articles in the BMJ from January to March 1976," 1 *Brit. Med. J*. 85 (1977).

[49] An-Wen Chan, Asbjørn Hróbjartsson, Mette T. Haahr, Peter C. Gøtzsche, and David G. Altman, D. G. "Empirical evidence for selective reporting of outcomes in randomized trials: Comparison of protocols to published articles," 291 *J. Am. Med. Ass'n* 2457 (2004).

[50] Wolfgang Ahrens & Iris Pigeot, eds., *Handbook of Epidemiology* 477 (2nd ed. 2014).

[51] Michael D. Green, D. Michal Freedman, and Leon Gordis, "Reference Guide on Epidemiology," in *Reference Manual on Scientific Evidence* 573 (3rd ed. 2011) 573 ("Study designs are developed before they begin gathering data.").

[52] John Bailar & Frederick Mosteller, "Guidelines for Statistical Reporting in Articles for Medical Journals," 108 *Ann. Intern. Med*. 2266, 268 (1988).

The necessity of a study protocol is "self-evident,"[53] and essential to research integrity.[54] The International Society of Pharmacoepidemiology has issued Guidelines for "Good Pharmacoepidemiology Practices,"[55] which calls for every study to have a written protocol. Among the requirements set out in this set of guidelines are descriptions of the research method, study design, operational definitions of exposure and outcome variables, and projected study sample size. The Guidelines provide that a detailed statistical analysis plan may be specified after data collection begins, but before any analysis commences.

Expert witness opinions on health effects are built upon studies, and so it behooves legal counsel to identify the methodological strengths and weaknesses of key studies through questioning whether they have protocols, whether the protocols were methodologically appropriate, and whether the researchers faithfully followed their protocols and their statistical analysis plans. Determining the peer review status of a publication, on the other hand, will often not advance a challenge based upon improvident methodology.

In some instances, a published study will have sufficiently detailed descriptions of methods and data that readers, even lawyers, can evaluate their scientific validity or reliability (*vel non*). In some cases, however, readers will be no better off than the peer reviewers who lacked access to protocols, statistical analysis plans, and original data. When a particular study is crucial support for an adversary's expert witness, a reasonable litigation goal may well be to obtain the protocol and statistical analysis plan, and if need be, the original underlying data. The decision to undertake such discovery is difficult. Discovery of non-party scientists can be expensive and protracted; it will almost certainly be contentious. When expert witnesses rely upon one or a few studies, which telegraph internal validity, this litigation strategy may provide the strongest evidence against the study's being reasonably relied upon, or its providing "sufficient facts and data" to support an admissible expert witness opinion.

V. Access to a Study Protocol and Underlying Data Reveals a Nuclear Non-Proliferation Test

The limits of peer review ultimately make it a poor proxy for the validity tests posed by Rules 702 and 703. Published peer review articles simply do not permit a very searching evaluation of the facts and data of a study. In the wake of the *Daubert* decision, expert witnesses quickly saw that they can obscure the search for validity by the reliance upon published studies, and frustrate the goals of judicial gatekeeping. As a practical matter, the burden shifts to the party that wishes to challenge the relied upon facts and data to learn more about the cited studies to show that the facts and data are not sufficient under Rule 702(b), and that the testimony is not the product of reliable methods under Rule 702(c). Obtaining study protocols, and in some instances, underlying data, is necessary for due process in the gatekeeping process. A couple of case studies

---

[53] Wolfgang Ahrens & Iris Pigeot, eds., *Handbook of Epidemiology* 477 (2nd ed. 2014).
[54] Sandra Alba, *et al*., "Bridging research integrity and global health epidemiology statement: guidelines for good epidemiological practice," 5 *BMJ Global Health* e003236, at p.3 & *passim* (2020).
[55] *See* "The ISPE Guidelines for Good Pharmacoepidemiology Practices (GPP)," available at <https://www.pharmacoepi.org/resources/policies/guidelines-08027/>.

may illustrate the power of looking under the hood of published studies, even ones that were peer reviewed.

When the Supreme Court decided the *Daubert* case in June 1993, two recent verdicts in silicone-gel breast implant cases were fresh in memory.[56] The verdicts were large by the standards of the time, and the evidence presented for the claims that silicone caused autoimmune disease was extremely weak. The verdicts set off a feeding frenzy, not only in the lawsuit industry, but also in the shady entrepreneurial world of supposed medical tests for "silicone sensitivity."

The plaintiffs' litigation theory lacked any meaningful epidemiologic support, and so there were fulsome presentations of putative, hypothetical mechanisms. One such mechanism involved the supposed *in vivo* degradation of silicone to silica (silicon dioxide), with silica then inducing an immunogenic reaction, which then, somehow, induced autoimmunity and the induction of autoimmune connective tissue disease. The degradation claim would ultimately prove baseless,[57] and the nuclear magnetic resonance evidence put forward to support degradation would turn out to be instrument artifact and deception. The immunogenic mechanism had a few lines of potential support, with the most prominent at the time coming from the laboratories of Douglas Radford Shanklin, and his colleague, David L. Smalley, both of whom were testifying expert witnesses for claimants.

The *Daubert* decision held out some opportunity to challenge the admissibility of testimony that silicone implants led to either the production of a silicone-specific antibody, or the induction of t-cell mediated immunogenicity from silicone (or resulting silica) exposure. The initial tests of the newly articulated standard for admissibility of opinion testimony in silicone litigation did not go well.[58] Peer review, which was absent in the re-analyses relied upon in the Bendectin litigation, was superficially present in the studies relied upon in the silicone litigation. The absence of

---

[56] Reuters, "Record $25 Million Awarded In Silicone-Gel Implants Case," *N.Y. Times* at A13 (Dec. 24, 1992) (describing the verdict returned in Harris County, Texas, in *Johnson v. Medical Engineering Corp*.); Associated Press, "Woman Wins Implant Suit," *N.Y. Times* at A16 (Dec. 17, 1991) (reporting a verdict in *Hopkins v. Dow Corning*, for $840,000 in compensatory and $6.5 million in punitive damages); *see Hopkins v. Dow Corning Corp*., 33 F.3d 1116 (9th Cir. 1994) (affirming judgment with minimal attention to Rule 702 issues).

[57] William E. Hull, "A Critical Review of MR Studies Concerning Silicone Breast Implants," 42 *Magnetic Resonance in Medicine* 984, 984 (1999) ("From my viewpoint as an analytical spectroscopist, the result of this exercise was disturbing and disappointing. In my judgement as a referee, none of the Garrido group's papers (1–6) should have been published in their current form."). *See also* N.A. Schachtman, "Silicone Data – Slippery & Hard to Find, Part 2," *Tortini* (July 5, 2015). Many of the material science claims in the breast implant litigation were as fraudulent as the health effects claims. *See, e.g.*, John Donley, "Examining the Expert," 49 *Litigation* 26 (Spring 2023) (discussing his encounters with frequent testifier Pierre Blais, in silicone litigation).

[58] *See, e.g.*, *Hopkins v. Dow Corning Corp*., 33 F.3d 1116 (9th Cir. 1994) (affirming judgment for plaintiff over Rule 702 challenges), *cert. denied*, 115 S.Ct. 734 (1995). *See* Donald A. Lawson, "Note, *Hopkins v. Dow Corning Corporation*: Silicone and Science," 37 *Jurimetrics J*. 53 (1996) (concluding that *Hopkins* was wrongly decided).

supportive epidemiology was excused with hand waving that there was a "credible" mechanism, and that epidemiology took too long and was too expensive. Initially, post-*Daubert*, federal courts were quick to excuse the absence of epidemiology for a novel claim.

The initial Rule 702 challenges to plaintiffs' expert witnesses thus focused on immunogenicity as the putative mechanism, which, if true, might lend some plausibility to their causal claim. Ultimately, plaintiffs' expert witnesses would have to show that the mechanism was real by showing that silicone exposure causes autoimmune disease through epidemiologic studies,

One of the more persistent purveyors of a "test" for detecting alleged silicone sensitivity came from Smalley and Shanklin, then at the University of Tennessee. These authors exploited the fears of implant recipients and the greed of lawyers by marketing a "silicone sensitivity test (SILS)." For a price, Smalley and Shanklin would test mailed-in blood specimens sent directly by lawyers or by physicians, and provide ready-for-litigation reports that claimants had suffered an immune system response to silicone exposure. Starting in 1995, Smalley and Shanklin also cranked out a series of articles at supposedly peer reviewed journals, which purported to identify a specific immune response to crystalline silica in women who had silicone gel breast implants.[59]

---

[59] *See* David L. Smalley, Douglas R. Shanklin, Mary F. Hall, and Michael V. Stevens, "Detection of Lymphocyte Stimulation by Silicon Dioxide," 4 *Internat'l J. Occup. Med. & Toxicol*. 63 (1995); David L. Smalley, Douglas R. Shanklin, Mary F. Hall, Michael V. Stevens, and Aram Hanissian, "Immunologic stimulation of T lymphocytes by silica after use of silicone mammary implants," 9 *FASEB J*. 424 (1995); David L. Smalley, J. J. Levine, Douglas R. Shanklin, Mary F. Hall, Michael V. Stevens, "Lymphocyte response to silica among offspring of silicone breast implant recipients," 196 *Immunobiology* 567 (1996); David L. Smalley, Douglas R. Shanklin, "T-cell-specific response to silicone gel," 98 *Plastic Reconstr. Surg*. 915 (1996); and Douglas R. Shanklin, David L. Smalley, Mary F. Hall, Michael V. Stevens, "T cell-mediated immune response to silica in silicone breast implant patients," 210 *Curr. Topics Microbiol. Immunol*. 227 (1996). Shanklin was also no stranger to making his case in the popular media. *See, e.g.*, Douglas Shanklin, "More Research Needed on Breast Implants," *Kitsap Sun* at 2 (Aug. 29, 1995) ("Widespread silicone sickness is very real in women with past and continuing exposure to silicone breast implants.") (writing for Scripps Howard News Service). Even after the Shanklin studies were discredited in court, Shanklin and his colleagues continued to publish their claims that silicone implants led to silica antigenicity. David L. Smalley, Douglas R. Shanklin, and Mary F. Hall, "Monocyte-dependent stimulation of human T cells by silicon dioxide," 66 *Pathobiology* 302 (1998); Douglas R. Shanklin and David L. Smalley, "The immunopathology of siliconosis. History, clinical presentation, and relation to silicosis and the chemistry of silicon and silicone," 18 *Immunol. Res*. 125 (1998); Douglas Radford Shanklin, David L. Smalley, "Pathogenetic and diagnostic aspects of siliconosis," 17 *Rev. Environ Health* 85 (2002), and "Erratum," 17 *Rev Environ Health*. 248 (2002); Douglas Radford Shanklin & David L Smalley, "Kinetics of T lymphocyte responses to persistent antigens," 80 *Exp. Mol. Pathol.* 26 (2006). Douglas Shanklin died in 2013. Susan J. Ainsworth, "Douglas R. Shanklin," 92 *Chem. & Eng'g News* (April 7, 2014). Dr. Smalley appears to be still alive. In 2022, he sued the federal government to challenge his disqualification from serving as a laboratory director of any clinical directory in the United States, under 42 U.S.C. § 263a(k). He lost. ***Smalley v. Becerra***, Case No. 4:22CV399 HEA (E.D. Mo. July 6, 2022).

These studies had two obvious goals. First, the studies promoted their product to the "silicone sisters," various support groups of claimants, as well as their lawyers, and a network of supporting rheumatologists and plastic surgeons. Second, by identifying a putative causal mechanism, Shanklin could add a meretricious patina of scientific validity to the claim that silicone breast implants cause autoimmune disease, which Shanklin, as a testifying expert witness, needed to survive Rule 702 challenges.

The plaintiffs' strategy was to paper over the huge analytical gaps in their causal theory with complicated, speculative research, which had been peer reviewed and published. Although the quality of the journals was often suspect, and the nature of the peer review obscure, the strategy was initially successful in deflecting any meaningful judicial scrutiny.

Many of the silicone cases were pending in a multi-district litigation, MDL 926, before Judge Sam Pointer, in the Northern District of Alabama. Judge Pointer, however, did not believe that ruling on expert witness admissibility was a function of an MDL court, and by 1995, he started to remand cases to the transferor courts, for those courts to do what they thought appropriate under Rules 702 and 703. Some of the first remanded cases went to the District of Oregon, where they landed in front of Judge Robert E. Jones. In early 1996, Judge Jones invited briefing on expert witness challenges, and in face of the complex immunology and toxicology issues, and the emerging epidemiologic studies, he decided to appoint four technical advisors to assist him in deciding the challenges.

The addition of scientific advisors to the gatekeeper's bench made a huge difference in the sophistication and detail of the challenges that could be lodged to the relied-upon studies. In June 1996, Judge Jones entertained extensive hearings with *viva voce* testimony from both challenged witnesses and subject-matter experts on topics, such as immunology, toxicology, epidemiology, and nuclear magnetic resonance spectroscopy. Judge Jones invited final argument in the form of videotaped presentations from counsel so that the videotapes could be distributed to his technical advisors later in the summer. The contrived complexity of plaintiffs' case dissipated, and the huge analytical gaps became visible. In December 1996, Judge Jones issued his decision that excluded the plaintiffs' expert witnesses' proposed testimony on grounds that it failed to satisfy the requirements of Rule 702.[60]

---

[60] *Hall v. Baxter Healthcare Corp.*, 947 F. Supp. 1387 (D. Ore. 1996); *see* Joseph Sanders & David H. Kaye, "Expert Advice on Silicone Implants: *Hall v. Baxter Healthcare Corp.*, 37 *Jurimetrics J.* 113 (1997); Laurens Walker & John Monahan, "Scientific Authority: The Breast Implant Litigation and Beyond," 86 *Virginia L. Rev.* 801 (2000); Jane F. Thorpe, Alvina M. Oelhafen, and Michael B. Arnold, "Court-Appointed Experts and Technical Advisors," 26 *Litigation* 31 (Summer 2000); Laural L. Hooper, Joe S. Cecil & Thomas E. Willging, "Assessing Causation in Breast Implant Litigation: The Role of Science Panels," 64 *Law & Contemp. Problems* 139 (2001); Debra L. Worthington, Merrie Jo Stallard, Joseph M. Price & Peter J. Goss, "Hindsight Bias, *Daubert*, and the Silicone Breast Implant Litigation: Making the Case for

In October 1996, while Judge Jones was studying the record and writing his opinion in the *Hall* case, Judge Weinstein, with a judge from the Southern District of New York, and another from New York state trial court, conducted a two-week Rule 702 hearing, in Brooklyn. Judge Weinstein announced at the outset that he had studied the record from the *Hall* case, and that he would incorporate it into his record for the cases remanded to the Southern and Eastern Districts of New York.

Curious gaps in the articles claiming silicone immunogenicity, and the lack of success in earlier Rule 702 challenges, motivated the defense to obtain the study protocols and underlying data from studies such as those published by Shanklin and Smalley. Shanklin and Smalley were frequently listed as expert witnesses in individual cases, but when requests or subpoenas for their protocols and raw data were filed, plaintiffs' counsel stonewalled or withdrew them as witnesses. Eventually, the defense was able to enforce a subpoena and obtain the protocol and some data. The respondents claimed that the control data no longer existed, and inexplicably a good part of the experimental data had been destroyed. Enough was revealed, however, to see that the published articles were not what they claimed to be.[61]

Court-Appointed Experts in Complex Medical and Scientific Litigation," 8 *Psychology, Public Policy & Law* 154 (2002).

[61] Judge Jones' technical advisor on immunology reported that the studies offered in support of the alleged connection between silicone implantation and silicone-specific T cell responses, including the published papers by Shanklin and Smalley, "have a number of methodological shortcomings and thus should not form the basis of such an opinion." Mary Stenzel-Poore, "Silicone Breast Implant Cases--Analysis of Scientific Reasoning and Methodology Regarding Immunological Studies" (Sept. 9, 1996). This judgment was seconded, over three years later, in the proceedings before MDL 926 and its Rule 706 court-appointed immunology expert witness. *See* Report of Dr. Betty A. Diamond, in MDL 926, at 14-15 (Nov. 30, 1998). Other expert witnesses who published studies on the supposed immunogenicity of silicone came up with some creative excuses to avoid producing their underlying data. Eric Gershwin consistently testified that his data were with a co-author in Israel, and that he could not produce them. N.A. Schachtman, "Silicone Data – Slippery and Hard to Find, Part I," Tortini (July 4, 2015). Nonetheless, the court-appointed technical advisors were highly critical of Dr. Gershwin's results. Dr. Stenzel-Poore, the immunologist on Judge Jones' panel of advisors, found Gershwin's claims "not well substantiated." *Hall v. Baxter Healthcare Corp*., 947 F.Supp. 1387 (D. Ore. 1996). Similarly, Judge Pointer's appointed expert immunologist Dr. Betty A. Diamond, was unshakeable in her criticisms of Gershwin's work and his conclusions. Testimony of Dr. Betty A. Diamond, in MDL 926 (April 23, 1999). And the Institute of Medicine committee, charged with reviewing the silicone claims, found Gershwin's work inadequate and insufficient to justify the extravagent claims that plaintiffs were making for immunogenicity and for causation of autoimmune disease. Stuart Bondurant, Virginia Ernster, and Roger Herdman, eds., *Safety of Silicone Breast Implants* 256 (1999). Another testifying expert witness who relied upon his own data, Nir Kossovsky, resorted to a seismic excuse; he claimed that the Northridge Quake destroyed his data. N.A. Schachtman, "Earthquake Induced Data Loss – We're All Shook Up," *Tortini* (June 26, 2015); Kossovsky, along with his wife, Beth Brandegee, and his father, Ram Kossowsky, sought to commercialize an ELISA-based silicone "antibody" biomarker

In addition to litigation discovery, in March 1996, a surgeon published the results of his test of the Shanklin-Smalley silicone sensitivity test ("SILS").[62] Dr. Leroy Young sent the Shanklin laboratory several blood samples from women with and without silicone implants. For six women who never had implants, Dr. Young submitted a fabricated medical history that included silicone implants and symptoms of "silicone-associated disease." All six samples were reported back as "positive"; indeed, these results were more positive than the blood samples from the women who actually had silicone implants. Dr. Young suggested that perhaps the SILS test was akin to cold fusion.

By the time counsel assembled in Judge Weinstein's courtroom, in October 1996, some additional epidemiologic studies had become available and much more information was available on the supposedly supportive mechanistic studies upon which plaintiffs' expert witnesses had previously relied. Not too surprisingly, plaintiffs' counsel chose not to call the entrepreneurial Dr. Shanklin, but instead called Donard S. Dwyer, an earnest, young immunologist who had done some contract work on an unrelated matter for Bristol-Myers Squibb, a defendant in the litigation. Dr. Dwyer had filed an affidavit previously in the Oregon federal litigation, in which he gave blanket approval to the methods and conclusions of the Smalley-Shanklin research:

> "Based on a thorough review of these extensive materials which are more than adequate to evaluate Dr. Smalley's test methodology, I formed the following conclusions. First, the experimental protocols that were used are standard and acceptable methods for measuring T Cell proliferation. The results have been reproducible and consistent in this laboratory. Second, the conclusion that there are differences between patients with breast implants and normal controls with respect to the proliferative response to silicon dioxide appears to be justified from the data."[63]

Dwyer maintained this position even after he reviewed the study protocol and underlying data, and the scathing evaluations of the Smalley-Shanklin work by the defense immunologists. On

---

diagnostic test, Detecsil. Although the early Rule 702 decisions declined to take a hard at Kossovsky's study, the U.S. Food and Drug Administration eventually shut down the Kossovsky Detecsil test. Lillian J. Gill, FDA Acting Director, Office of Compliance, Letter to Beth S. Brandegee, President, Structured Biologicals (SBI) Laboratories: Detecsil Silicone Sensitivity Test (July 15, 1994); *see* Gary Taubes, "Silicone in the System: Has Nir Kossovsky really shown anything about the dangers of breast implants?" *Discover Magazine (Dec. 1995).*

[62] Leroy Young, "Testing the Test: An Analysis of the Reliability of the Silicone Sensitivity Test (SILS) in Detecting Immune-Mediated Responses to Silicone Breast Implants," 97 *Plastic & Reconstr. Surg*. 681 (1996).

[63] Affid. of Donard S. Dwyer, at para. 6 (Dec. 1, 1995), filed in *In re Breast Implant Litig. Pending in U.S. D. Ct, D. Oregon* (Groups 1,2, and 3).

direct examination at the hearings in Brooklyn, Dwyer vouched for the challenged t-cell studies, and opined that the work was peer reviewed and sufficiently reliable.[64]

The charade fell apart on cross-examination. Dwyer refused to endorse the studies that claimed to have found an anti-silicone antibody. Researchers at leading universities had attempted to reproduce the findings of such antibodies, without success.[65] The real controversy was over the claimed finding of silicone antigenicity as shown in t-cell or other cell-mediated specific immune response. On direct examination, plaintiffs' counsel elicited Dwyer's support for the soundness of the scientific studies that purported to establish such antigenicity, with little attention to the critiques that had been filed before the hearing.[66] Dwyer stuck to his unqualified support he had expressed previously in his affidavit for the Oregon cases.[67]

The problematic aspect of Dwyer's direct examination testimony was that he had seen the protocol and the partial data produced by Smalley and Shanklin.[68] Dwyer, therefore, could not resist some basic facts about their work. First, the Shanklin data failed to support a dose-response relationship.[69] Second, the blood samples from women with silicone implants had been mailed to Smalley's laboratory, whereas the control samples were collected locally. The disparity ensured that the silicone blood samples would be older than the controls, which was a departure from treating exposed and control samples in the same way.[70] Third, the experiment was done unblinded; the laboratory technical personnel and the investigators knew which blood samples were silicone exposed and which were controls (except for samples sent by Dr. Leroy Young).[71] Fourth, Shanklin's laboratory procedures deviated from the standardized procedure set out in the National Institute of Health's *Current Protocols in Immunology*.[72]

The SILS study protocol and the data produced by Shanklin and Smalley made clear that each sample was to be tested in triplicate for t-cell proliferation in response to silica, to a positive control mitogen (Con A), and to a negative control blank. The published papers all claimed that the each sample was tested in triplicate for each of these three response situations (silica, mitogen, and nothing).[73] Shanklin and Smalley described their t-cell proliferation studies, in their published papers, as having been done in triplicate. These statements were, however, untrue and never corrected.[74]

---

[64] Notes of Testimony of Dr. Donnard Dwyer, *Nyitray v. Baxter Healthcare Corp*., CV 93-159 (E. & S.D.N.Y and N.Y. Sup. Ct., N.Y. Cty. Oct. 8, 9, 1996) (Weinstein, J., Baer, J., Lobis, J., Pollak, M.J.).
[65] *Id*. at N.T. 238-239 (Oct. 8, 1996).
[66] *Id*. at N.T. 240.
[67] *Id*. at N.T. 241-42.
[68] *Id*. at N.T. 243-44; 255:22-256:3.
[69] *Id*. at 244-45.
[70] *Id*. at N.T. 259.
[71] *Id*. at N.T. 258:20-22.
[72] *Id*. at N.T. 254.
[73] *Id*. at N.T. 252:16-254.
[74] *Id*. at N.T. 254:19-255:2.

The study protocol called for the tests to be run in triplicate, but they instructed the laboratory that two counts may be used if one count does not match the other counts, which was to be decided by a technical specialist on a "case-by-case" basis. Of data that was supposed to be reported in triplicate, fully one third had only two data points, and 10 percent had but one data point.[75] No criteria were provided to the technical specialist for deciding which data to discard.[76] Not only had Shanklin excluded data, but he discarded and destroyed the data such that no one could go back and assess whether the data should have been excluded.[77]

Dwyer agreed that this exclusion and discarding of data was not at all a good method.[78] Dwyer proclaimed that he had not come to Brooklyn to defend this aspect of the Shanklin work, and that it was not defensible at all. Dwyer conceded that "the interpretation of the data and collection of the data are flawed."[79] Dwyer tried to stake out a position that was incoherent by asserting that there was "nothing inherently wrong with the method," while conceding that discarding data was problematic.[80] The judges presiding over the hearing could readily see that the Shanklin research was bent.

At this point, the lead plaintiffs' counsel, Michael Williams, sought an off-ramp. He jumped to his feet and exclaimed "I'm informed that no witness in this case will rely on Dr. Smalley's [and Shanklin's] work in any respect." [81] Judge Weinstein's eyes lit up with the prospect that the Smalley-Shanklin work, by agreement, would never be mentioned again in New York state or federal cases. Given how central the claim of silicone antigenicity was to plaintiffs' cases, the defense resisted the stipulation about research that they would continue to face in other state and federal courts. The defense was saved, however, by the obstinance of a lawyer from the Weitz & Luxenberg firm, who rose to report that her firm intended to call Drs. Shanklin and Smalley as witnesses, and that they would not stipulate to the exclusion of their work. Judge Weinstein rolled his eyes, and waved the defense examiner to continue.[82] The proliferation of t-cell tests was over. The hearing before Judges Weinstein and Baer, and Justice Lobis, continued for several more days, with several other dramatic moments.[83]

In short order, on October 23, 1996, Judge Weinstein issued a short, published opinion, in which he granted partial summary judgment on the claims of systemic disease for all cases pending in

[75] *Id*. at N.T. 269:18-269:14.
[76] *Id*. at N.T. 261:23-262:1.
[77] *Id*. at N.T. 269:18-270.
[78] *Id*. atN.T. 256:3-16.
[79] *Id*. at N.T. 262:15-17
[80] *Id*. at N.T. 247:3-5.
[81] *Id*. at N.T. at 260:2-3
[82] *Id*. at N.T. at 261:5-8.
[83] One of the more interesting and colorful moments came when James Conlon cross-examined plaintiffs' pathology expert witness, Saul Puszkin, about questionable aspects of his curriculum vitae. The examination was revealed such questionable conduct that Judge Weinstein stopped the examination and directed Dr. Puszkin not to continue without legal counsel of his own.

federal court in New York.[84] What was curious was that the defendants had ***not*** moved for summary judgment. There were, of course, pending motions to exclude plaintiffs' expert witnesses, but Judge Weinstein effectively ducked those motions, and let it be known that he was never a fan of Rule 702. It would be many years later before Judge Weinstein allowed his judicial assessment see the light of day. Two decades and some years later, in a law review article, Judge Weinstein gave his judgment that

> "[t]he breast implant litigation was largely based on a litigation fraud. … Claims—supported by **medical charlatans**—that enormous damages to women's systems resulted could not be supported."[85]

Judge Weinstein's opinion was truly a judgment from which there could be no appeal. Shanklin and Smalley continued to publish papers for another decade. None of the published articles by Shanklin and others have been retracted.

VI. How Access to a Protocol and Underlying Data Gave Yale Researchers a Big Black Eye

A. Prelude to Litigation

Phenylpropanolamine (PPA) was a widely used direct α-adrenergic agonist used as a medication to control cold symptoms and to suppress appetite for weight loss.[86] In 1972, an over-the-counter (OTC) Advisory Review Panel, to the U.S. Food and Drug Administration (FDA) considered the safety and efficacy of PPA-containing nasal decongestant medications, leading, in 1976, to a recommendation that the agency label these medications as "generally recognized as safe and effective." Several years later, in 1982, another FDA Panel recommended that PPA-containing weight control products also be recognized as safe and effective.

Two epidemiologic studies of PPA and hemorrhagic stroke (HS) were conducted in the 1980s. The results of one study by Hershel Jick and colleagues, presented as a letter to the editor, reported a relative risk of 0.58, with a 95% exact confidence interval, 0.03 - 2.9.[87] A year later, two researchers, reporting a study based upon Medicaid databases, found no significant associations between PPA use and HS.[88]

---

[84] *In re Breast Implant Cases*, 942 F. Supp. 958 (E.& S.D.N.Y. 1996). The opinion did not specifically address the Rule 702 and 703 issues that were the subject of pending motions before the court.

[85] Hon. Jack B. Weinstein, "Preliminary Reflections on Administration of Complex Litigation" 2009 *Cardozo L. Rev. de novo* 1, 14 (2009) (emphasis added).

[86] Rachel Gorodetsky, "Phenylpropanolamine," in Philip Wexler, ed., 7 *Encyclopedia of Toxicology* 559 (4th ed. 2024).

[8787] Hershel Jick, Pamela Aselton, and Judith R. Hunter, "Phenylpropanolamine and Cerebral Hemorrhage," 323 *Lancet* 1017 (1984).

[88] Robert R. O'Neill & Stephen W. Van de Carr, "A Case-Control Study of Adrenergic Decongestants and Hemorrhagic CVA Using a Medicaid Data Base" m.s. (1985).

The FDA, however, did not approve a final monograph for PPA, with recognition of its "safe and effective" status because of occasional reports of HS that occurred in patients who used PPA-containing medications, mostly young women who had used PPA appetite suppressants for dieting. In 1982, the FDA requested information on the effects of PPA on blood pressure, particularly with respect to weight-loss medications. The agency deferred a proposed 1985 final monograph because of the blood pressure issue.

The FDA deemed the data inadequate to answer its safety concerns. Congressional and agency hearings in the early 1990s amplified some public concern, but in 1990, the Director of Cardio-Renal Drug Products, at the Center for Drug Evaluation and Research, found several well-supported facts, based upon robust evidence. Blood pressure studies in humans showed a biphasic response. PPA initially causes blood pressure to rise above baseline (a pressor effect), and then to fall below baseline (depressor effect). These blood pressure responses are dose-related, and diminish with repeated use. Patients develop tolerance to the pressor effects within a few hours. The Center concluded that at doses of 50 mg of PPA and below, the pressor effects of the medication are small, indeed smaller than normal daily variations in basal blood pressure. Humans develop tolerance to the pressor effects quickly, within the time frame of a single dose. The only time period in which even a theoretical risk might exist is within a few hours, or less, of a patient's taking the first dose of PPA medication. Doses of 25 mg. immediate-release PPA could not realistically be considered to pose any "absolute safety risk and have a reasonable safety margin."[89]

In 1991, Dr. Heidi Jolson, an FDA scientist wrote that the agency's spontaneous adverse event reporting system "suggested" that PPA appetite suppressants increased the risk of cerebrovascular accidents. A review of stroke data, including the adverse event reports, by epidemiology consultants failed to support a causal association between PPA and hemorrhagic stroke (HS). The reviewers, however, acknowledged that the available data did not permit them to rule out a risk of HS. The FDA adopted the reviewers' recommendation for a prospective, large case-control study designed to take into account the known physiological effects of PPA on blood pressure.[90]

What emerged from this regulatory indecision was a decision to conduct another epidemiologic study. In November 1992, a manufacturers' group, now known as the Consumer Healthcare Products Association (CHPA), proposed a case-control study that would become known as the Hemorrhagic Stroke Project (HSP). In March 1993, the group submitted a proposed protocol, and a suggestion that the study be conducted by several researchers at Yale University. After feedback from the public and the Yale researchers, the group submitted a final protocol in April 1994. Both the researchers and the sponsors agreed to a scientific advisory group that would operate independently and oversee the study. The study began in September 1994. The FDA deferred action on a final monograph for PPA, and product marketing continued.

---

[89] Ramond Lipicky, Center for Drug Evaluation and Research, PPA, Safety Summary at 29 (Aug. 9, 1900).

[90] Center for Drug Evaluation and Research, US Food and Drug Administration, "Epidemiologic Review of Phenylpropanolamine Safety Issues" (April 30, 1991).

The Yale HSP authors delivered their final report on their case-control study to FDA, in May 2000.[91] The HSP was a study, with 702 HS cases, and over 1,376 controls, men and women, ages 18 to 49. The report authors concluded that "the results of the HSP *suggest* that PPA increases the risk for hemorrhagic stroke."[92] The study had taken over five years to design, conduct, and analyze. In September 2000, the FDA's Office of Post-Marketing Drug Risk Assessment released the results, with its own interpretation and conclusion that dramatically exceeded the HSP authors' own interpretation.[93] The FDA's Non-Prescription Drug Advisory Committee then voted, on October 19, 2000, to recommend that PPA be reclassified as "unsafe." The Committee's meeting, however, was attended by several leading epidemiologists who pointed to important methodological problems and limitations in the design and execution of the HSP.[94]

In November 2000, the FDA" Nonprescription Drugs Advisory Committee determined that there was a significant association PPA and HS, and recommended that PPA not be considered safe for OTC use. The FDA never addressed causality; nor did it have to do so under governing law. The FDA's actions led the drug companies voluntarily to withdraw PPA-containing products.

B. The Hemorrhagic Stroke Project

The December 21, 2000, issue of *The New England Journal of Medicine* featured a revised version of the HSP report as its lead article.[95] Under the journal's guidelines for statistical reporting, the authors were required to present two-tailed p-values or confidence intervals. Results from the HSP Final Report looked considerably less impressive after the obtained significance probabilities were doubled. Only the finding in appetite suppressant use was branded an independent risk factor:

---

[91] Ralph I. Horwitz, Lawrence M. Brass, Walter N. Kernan, Catherine M. Viscoli, "Phenylpropanolamine & Risk of Hemorrhagic Stroke – Final Report of the Hemorrhagic Stroke Project (May 10, 2000).

[92] *Id*. at 3, 26 (emphasis added).

[93] Lois La Grenade & Parivash Nourjah, "Review of study protocol, final study report and raw data regarding the incidence of hemorrhagic stroke associated with the use of phenylopropanolamine," Division of Drug Risk Assessment, Office of Post-Marketing Drug Risk Assessment (0PDRA) (Sept. 27, 2000). These authors concluded that the HSP report provided "compelling evidence of increased risk of hemorrhagic stroke in young people who use PPA-containing appetite suppressants. This finding, taken in association with evidence provided by spontaneous reports and case reports published in the
medical literature leads us to recommend that these products should no longer be available for over the counter use."

[94] Among those who voiced criticisms of the design, methods, and interpretation of the HSP study were Noel Weiss, Lewis Kuller, Brian Strom, and Janet Daling. Many of the criticisms would prove to be understated in the light of post-publication review.

[95] Walter N. Kernan, Catherine M. Viscoli, Lawrence M. Brass, J.P. Broderick, T. Brott, and Edward Feldmann, "Phenylpropanolamine and the risk of hemorrhagic stroke," 343 *New Engl. J. Med*. 1826 (2000) [cited as Kernan]

> "The results suggest that phenylpropanolamine in appetite suppressants, and *possibly* in cough and cold remedies, is an independent risk factor for hemorrhagic stroke in women."[96]

The HSP had multiple pre-specified aims, and several other statistical comparisons and analyses were added along the way. No statistical adjustment was made for these multiple comparisons, but their presence in the study must be considered. Perhaps that is why the authors merely suggest that PPA in appetite suppressants was an independent risk factor for HS in women. Under current statistical guidelines for the *New England Journal of Medicine*, this suggestion might require even further qualification and weakening.[97]

The HSP study faced difficult methodological issues. The detailed and robust identification of PPA's blood pressure effects in humans focused attention on the crucial timing of HS in relation to ingestion of a PPA medication. Any use, or any use within the last seven or 30 days, would be fairly irrelevant to the pathophysiology of cerebral hemorrhage, given the known physiological effects and pharmokinetics of PPA. The HSP authors settled on a definition of "first use" as any use of a PPA product within 24 hours, and no other uses in the previous two weeks.[98] Given the rapid onset of pressor and depressor effects, and adaptation response, this definition of first use was generous and likely included many irrelevant exposed cases, but at least the definition attempted to incorporate the phenomena of short-lived effect and adaption. The appetite suppressant association did not involve any "first use," which makes the one "suggested" increase risk seem much less certain and relevant.

The alternative definition of exposure, in addition to "first use," the ingestion of the PPA-containing medication took place as "the index day before the focal time and the preceding three calendar days." Again, given the known pharmacokinetics and physiological effects of PPA, this three-day (plus) window seems doubtfully relevant.

All instances of "first use" occurred among men and women who used a cough or cold remedy, with an adjusted OR of 3.14, with a 95% confidence interval (CI), of 0.96–10.28), p = 0.06. The very wide confidence interval, in excess of an order of magnitude, reveals the fragility of the statistical inference. There were but 8 first use exposed stroke cases (out of 702), and 5 exposed controls (out of 1,376).

When this first use analysis is broken down between men and women, the result becomes even more fragile. Among men, there was only one first use exposure in 319 male HS patients, and one first use exposure in 626 controls, for an adjusted OR of 2.95, CI 0.15 - 59.59, and p = 0.48. Among women, there were 7 first use exposures among 383 female HS patients, and 4 first use exposures among 750 controls, with an adjusted OR of 3.13, CI 0.86 - 11.46, p = 0.08.

---

[96] *Id*. at 1826 (emphasis added).
[97] David Harrington, Ralph B. D'Agostino, Sr., Constantine Gatsonis, Joseph W. Hogan, David J. Hunter, Sharon-Lise T. Normand, Jeffrey M. Drazen, and Mary Beth Hamel, "New Guidelines for Statistical Reporting in the Journal," 381 *New Engl. J. Med*. 285 (2019).
[98] Kernan, *supra* note 95, at 1827.

The small numbers of actual first exposure cases speak loudly for the inconclusiveness and fragility of the study results, and the sensitivity of the results to any methodological deviations or irregularities. Of course, for the one "suggested" association for appetite suppressant use among women, the results were even more fragile. None of the appetite suppressant cases were "first use," which raises serious questions whether anything meaningful was measured. There were six (non-first use) exposed among 383 female HS patients, with only a single exposed female control among 750. The authors presented an adjusted OR of 15.58, with a p-value of 0.02. The CI, however, spanned more than two orders of magnitude, 1.51 – 182.21, which makes the result well-nigh uninterpretable. One of six appetite suppressant cases was also a user of cough-cold remedies, and she was double counted in the study's analyses. This double-counted case, had a body-mass index of 19, which is certainly not overweight, and at the low end of normal.[99] The one appetite suppressant control was obese.

For the more expansive any-exposure analysis for use of PPA cough-cold medication, the results were significantly unimpressive. There were six exposed male cases among 391 male HS cases, and 13 exposed controls, for an adjusted odds ratio of 0.62, CI 0.20 – 1.92, p = 0.41. Although not an inverse association, the sample results for men were incompatible with a hypothetical doubling of risk. For women, on the expansive exposure definition, there were 16 exposed cases, among 383 female cases, with 19 exposed controls out of 750 female controls. The odds ratio for female PPA cough-cold medication was 1.54, CI 0.76 - 3.14, p = 0.23.

Aside from doubts whether the HSP measured meaningful exposures, the small number of exposed cases and controls present insuperable interpretative difficulties for the study. First, working with a case-control design and odds ratios, there should be some acknowledgment that odds ratios always exaggerate the observed association size compared with a relative risk.[100] Second, the authors knew that confounding would be an important consideration in evaluating any observed association. Known and suspected risk factors were consistently more prevalent among cases than controls.[101]

The HSP authors valiantly attempted to control for confounding in two ways. They selected controls by a technique known as random digit dialing, to find two controls for each case, matched on telephone exchange, sex, age, and race. The HSP authors, however, used imperfectly matched controls rather than lose the corresponding case from their study.[102] For other co-variates, the authors used multivariate logistic regression to provide odds ratios that were

---

[99] Transcript of Meeting on Safety Issues of Phenylpropanolamine (PPA) in Over-the-Counter Drug Products 117 (Oct. 19, 2000).

[100] *See, e.g.*, Huw Talfryn Oakley Davies, Iain Kinloch Crombie, and Manouche Tavakoli, "When can odds ratios mislead?" 316 *Brit. Med. J*. 989 (1998); Thomas F. Monaghan, Rahman, Christina W. Agudelo, Alan J. Wein, Jason M. Lazar, Karel Everaert, and Roger R. Dmochowski, "Foundational Statistical Principles in Medical Research: A Tutorial on Odds Ratios, Relative Risk, Absolute Risk, and Number Needed to Treat," 18 *Internat'l J. Envt'l Research & Public Health* 5669 (2021).

[101] Kernan, *supra* note 95, at 1829, Table 2.

[102] *Id*, at 1827.

adjusted for potential confounding from the measured covariates. At least two covariates, alcohol and cocaine use, involved potential legal or moral judgment, which almost certainly would have skewed HSP interview results.

An even more important threat to methodological validity, key covariates, such as smoking, alcohol use, hypertension, and cocaine use, were incorporated into the adjustment regression as dichotomous variables; body mass index was entered as a polychotomous variable. These factors are typically measured as continuous variables, and operate as potential confounders in proportion to continuous measures. Monte Carlo simulation shows that categorizing a continuous variable in logistic regression results in inflating the rate of finding false positive associations.[103] The type I (false-positive) error rates increases with sample size, with increasing correlation between the confounding variable and outcome of interest, and the number of categories used for the continuous variables. Numerous authors have warned of the cost and danger of dichotomizing continuous variables, in losing information, statistical power, and reliability.[104] In the field of pharmaco-epidemiology, the bias created by dichotomization of a continuous variable is harmful from both the perspective of statistical estimation and hypothesis testing.[105] Readers will be misled into believing that a study has adjusted for important covariates with the false allure of fully adjusted model.

Finally, with respect to the use of logistic regression to control confounding and provide adjusted odds ratios, there is the problem of the small number of events. Although the overall sample size

---

[103] Peter C. Austin & Lawrence J. Brunner, "Inflation of the type I error rate when a continuous confounding variable is categorized in logistic regression analyses," 23 *Statist. Med*. 1159 (2004).

[104] *See, e.g.*, Douglas G. Altman & Patrick Royston, "The cost of dichotomising continuous variables," 332 *Brit. Med. J*. 1080 (2006); Patrick Royston, Douglas G. Altman, and Willi Sauerbrei, "Dichotomizing continuous predictors in multiple regression: a bad idea," 25 *Stat. Med*. 127 (2006). *See also* Robert C. MacCallum, Shaobo Zhang, Kristopher J. Preacher, and Derek D. Rucker, "On the Practice of Dichotomization of Quantitative Variables," 7 *Psychological Methods* 19 (2002); David L. Streiner, "Breaking Up is Hard to Do: The Heartbreak of Dichotomizing Continuous Data," 47 *Can. J. Psychiatry* 262 (2002); Henian Chen, Patricia Cohen, and Sophie Chen, "Biased odds ratios from dichotomization of age," 26 *Statist. Med*. 3487 (2007); Carl van Walraven & Robert G. Hart, "Leave 'em Alone – Why Continuous Variables Should Be Analyzed as Such," 30 *Neuroepidemiology* 138 (2008); O. Naggara, J. Raymond, F. Guilbert, D. Roy, A. Weill, and Douglas G. Altman, "Analysis by Categorizing or Dichotomizing Continuous Variables Is Inadvisable," 32 *Am. J. Neuroradiol*. 437 (Mar 2011); Neal V. Dawson & Robert Weiss, "Dichotomizing Continuous Variables in Statistical Analysis: A Practice to Avoid," *Med. Decision Making* 225 (2012); Phillippa M Cumberland, Gabriela Czanner, Catey Bunce, Caroline J Doré, Nick Freemantle, and Marta García-Fiñana, "Ophthalmic statistics note: the perils of dichotomising continuous variables," 98 *Brit. J. Ophthalmol*. 841 (2014).

[105] Valerii Fedorov, Frank Mannino1, and Rongmei Zhang, "Consequences of dichotomization," 8 *Pharmaceut. Statist*. 50 (2009).

is adequate for logistic regression, cell sizes of one, or two, or three, raise serious questions about the use of large-sample statistical methods for analysis of the HSP results.[106]

C. Surfeit of Sub-Groups

The study protocol identified three (really four or five) specific goals, to estimate the associations: (1) between PPA use and HS; (2) between HS and type of PPA use (cough-cold remedy or appetite suppression); and (3) in women, between PPA appetite suppressant use and HS, and between PPA first use and HS.[107]

With two different definitions of "exposure," and some modifications added along the way, with two sexes, two different indications (cold remedy and appetite suppression), and with non-pre-specified analyses such as men's cough-cold PPA use, there was ample opportunity to inflate the Type I error rate. As the authors of the HSP final report acknowledged, they were able to identify only 60 "exposed" cases and controls.[108] In the context of a large case-controls study, the authors were able to identify some nominally statistically significant outcomes (PPA appetite suppressant and HS), but these were based upon very small numbers (six and one exposed, cases and controls, respectively), which made the results very uncertain considering the potential biases and confounding.

D. Design and Implementation Problems

Case-control studies always present some difficulty of obtaining controls that are similar to cases except that they did not experience the outcome of interest. As noted, controls were selected using "random digit dialing" in the same area code as the cases. The investigators were troubled by poor response rates from potential controls. They deviated from standard methodology for enrolling controls through random digit dialing by enrolling the first eligible control who agreed to participate, while failing to call back candidates who had asked to speak at another time.[109]

The exposure prevalence rate among controls was considerably lower than shown from PPA-product marketing research. This again raises questions about the low reported exposure rates among controls, which would inflate any observed odds ratios. Of course, it seems eminently reasonable to predict that persons who were suffering from head colds or the flu might not answer their phones or might request a call back. People who are obese might be reluctant to tell a stranger on the telephone that they are using a medication to suppress their appetite.

---

[106] Peter Peduzzi, John Concato, Elizabeth Kemper, Theodore R. Holford, and Alvan R. Feinstein, "A simulation study of the number of events per variable in logistic regression analysis?" 49 J. *Clin. Epidem.* 1373 (1996).
[107] HSP Final Report at 5.
[108] HSP Final Report at 26.
[109] Byron G. Stier & Charles H. Hennekens, "Phenylpropanolamine and Hemorrhagic Stroke in the Hemorrhagic Stroke Project: A Reappraisal in the Context of Science, the Food and Drug Administration, and the Law," 16 *Ann. Epidem.* 49, 50 (2006) [cited as Stier & Hennekens].

In the face of this obvious opportunity for selection bias, there was also ample room for recall bias. Cases were asked about medication use just before an unforgettable catastrophic event in their lives. Controls were asked about medication use before a day within the range of the previous week. More controls were interviewed by phone than were cases. Given the small number of exposed cases and controls, recall bias created by the differential circumstances and interview settings and procedures, was never excluded.

E. Lumpen Epidemiology ICH vs SAH

Every epidemiologic study or clinical trial has an exposure and outcome of interest, in a population of interest. The point is to compare exposed and unexposed persons, of relevant age, gender, and background, with comparable risk factors other than the exposure of interest, to determine if the exposure makes any difference in the rate of events of the outcome of interest.

Composite end points represent "lumping" together different individual end points for consideration as a single outcome. The validity of composite end points depends upon assumptions, which will have to be made at the time investigators design their study and write their protocol. After the data are collected and analyzed, the assumptions may or may not be supported.

Lumping may offer some methodological benefits, such as increasing statistical power or reducing sample size requirements. Standard epidemiologic practice, however, as reflected in numerous textbooks and methodology articles, requires the reporting of the individual constitutive end points, along with the composite result. Even when the composite end point was employed based upon a view that the component end points are sufficiently related, that view must itself ultimately be tested by showing that the individual end points are, in fact, concordant, with risk ratios in the same direction.

There are many clear statements that caution the consumers of medical studies against being misled by misleading claims that may be based upon composite end points, in the medical literature. In 2004, the *British Medical Journal* published a useful paper, "Users' guide to detecting misleading claims in clinical research reports." One of the authors' suggestions to readers was:

> "Beware of composite endpoints."[110]

The one methodological point to which virtually all writers agree is that authors should report the results for the composite end point separately to permit readers to evaluate the individual results.[111] A leading biostatistical methodologist, the late Douglas Altman, cautioned readers

---

[110] Victor M. Montori, Roman Jaeschke, Holger J. Schünemann, Mohit Bhandari, Jan L Brozek, P. J. Devereaux, and Gordon H. Guyatt, "Users' guide to detecting misleading claims in clinical research reports," 329 *Brit. Med. J.* 1093 (2004).

[111] Wolfgang Ahrens & Iris Pigeot, eds., *Handbook of Epidemiology* 1840 (2d ed. 2014) (47.5.8 Use of Composite Endpoints); Stuart J. Pocock, John J. V. McMurray, and Tim J. Collier, "Statistical Controversies in Reporting of Clinical Trials: Part 2 of a 4-Part Series on Statistics

against assuming that the overall estimate of association can be interpreted for each individual end point, and advised authors to provide "[a] clear listing of the individual endpoints and the number of participants experiencing them" to permit a more meaningful interpretation of composite outcomes.[112]

The HSP authors used a composite of hemorrhagic strokes, which was composed of both intracerebral hemorrhages (ICH) and subarachnoid hemorrhages (SAH). In their *New England Journal of Medicine* article, the authors presented the composite end point, but not the risk ratios for the two individual end points. Before they published the article, one of the authors wrote his fellow authors to advise them that because ICH and SAH are very different medical phenomena, they should present the individual end points in their analysis.[113]

The HSP researchers eventually did publish an analysis of SAH and PPA use, in a separate paper.[114] The authors identified 425 SAH cases, of which 312 met the criteria for aneurysmal SAH. They looked at many potential risk factors such as smoking (OR = 5.07), family history (OR = 3.1), marijuana (OR = 2.38), cocaine (OR = 24.97), hypertension (OR = 2.39), aspirin (OR = 1.24), alcohol (OR = 2.95), education, as well as PPA.

Only a bivariate analysis was presented for PPA, with an odds ratio of 1.15, p = 0.87. No confidence intervals were presented. The authors were a bit more forthcoming about the potential role of bias and confounding in this publication than they were in their earlier 2000 HSP paper. "Biases that might have affected this analysis of the HSP include selection and recall bias."[115] When the misleading aspect of the composite came under attack in litigation, the federal judge assessing the validity of proffered opinions rejected the call to look at ICH and SAH separately and criticized the defense argument "this article demonstrates the lack of an association between PPA and SAHs resulting from the rupture of an aneurysm."[116]

If, as the court reports, the defendants actually claimed a "demonstration" of "the lack of association," then shame, and more shame, on them! There was good reason, however, to assert that there was no PPA-SAH association shown. First, the cited study provided only a bivariate analysis for PPA and SAH. The odds ratio of 1.15 pales in comparison the risk ratios reported for many other common exposures. We can only speculate what happens to the 1.15, when the PPA

---

for Clinical Trials," 66 *J. Am. Coll. Cardiol*. 2648, 2650-51 (2015) ("Interpret composite endpoints carefully."); Schulz & Grimes, "Multiplicity in randomized trials I: endpoints and treatments," 365 *Lancet* 1591, 1595 (2005).

[112] Eric Lim, Adam Brown, Adel Helmy, Shafi Mussa & Douglas Altman, "Composite Outcomes in Cardiovascular Research: A Survey of Randomized Trials," 149 *Ann. Intern. Med*. 612 (2008).

[113] *See, e.g*., Thomas Brott email to Walter Kernan (Sept. 10, 2000).

[114] Joseph P. Broderick, Catherine M. Viscoli, Thomas Brott, Walter N. Kernan, Lawrence M. Brass, Edward Feldmann, Lewis B. Morgenstern, Janet Lee Wilterdink, and Ralph I. Horwitz, "Major Risk Factors for Aneurysmal Subarachnoid Hemorrhage in the Young Are Modifiable," 34 *Stroke* 1375 (2003).

[115] *Id*. at 1379.

[116] *Id*. at 1243.

exposure is placed in a fully adjusted model for all important covariates. Second, the p-value of 0.87 does not tell that 1.15 is unreal or due to chance. The HSP reported a 15% increase in odds ratio, *which is very compatible with no risk at all*. Perhaps if the defendants had been more modest in their characterization they would not have given the court the basis to find that "defendants distort and misinterpret the *Stroke* Article."[117]

Rejecting the defendants' characterization, the court drew upon an affidavit from plaintiffs' expert witness, Kenneth Rothman, who explained that a p-value cannot provide evidence of lack of an effect.[118] A high p-value, with its corresponding 95% confidence interval that includes 1.0, can, however, show that the sample data are compatible with the null hypothesis. What the reviewing court missed, and the defendants may not have said effectively, is that the statistical analysis was a test of an hypothesis, and the test failed to allow for the rejection of the null hypothesis. The HSP's separate SAH paper left the status of PPA at best as an indeterminant, from which there could be no valid inference of an association between PPA use and aneurismal SAH.

F. I Once Was Blind, But Now I See

The HSP protocol called for interviewers to be blinded to the study hypothesis, but this guard against bias was abandoned.[119] The HSP report acknowledged that "[b]linding would have provided extra protection against unequal ascertainment of PPA exposure in case subjects compared with control subjects."[120]

The study was conducted out of four sites, and at least one of the sites violated protocol by informing cases that they were participating in a study designed to evaluate PPA and HS.[121] The published article in the *New England Journal of Medicine* misleadingly claimed that study participants were blinded to its research hypotheses.[122] Although the plaintiffs' expert witnesses tried to slough off this criticism, the lack of blinding among interviewers and study subjects amplifies recall biases, especially when study subjects and interviewers may have been reluctant to discuss fully several of the covariate exposures, such as cocaine, marijuana, and alcohol use.[123]

G. No Causation At All

---

[117] *Id*. at 1243.
[118] *Id*., citing Rothman Affidavit, ¶ 7; Kenneth J. Rothman, *Epidemiology: An Introduction* at 117 (2002).
[119] HSP Final Report at 26 (''HSP interviewers were not blinded to the case-control status of study subjects and some were aware of the study purpose'."); Walter Kernan Dep. at 473-74, *In re PPA Prods. Liab. Litig*., MDL 1407 (W.D. Wash.) (Sept. 19, 2002).
[120] HSP Final Report at 26.
[121] Stier & Hennekens, note 109 *supra*, at 51.
[122] NEJM at 1831.
[123] *See* Christopher T. Robertson & Aaron S. Kesselheim, *Blinding as a Solution to Bias – Strengthening Biomedical Science, Forensic Science, and the Law* 53 (2016); Sandy Zabell, "The Virtues of Being Blind," 29 *Chance* 32 (2016).

Scientists and the general population alike have been conditioned to view the controversy over tobacco smoking and lung cancer as a contrivance of the tobacco industry. What is lost in this conditioning is the context of Sir Arthur Bradford Hill's triumphant 1965 Royal Society of Medicine presidential address. Hill, along with his colleague Sir Richard Doll, were not overly concerned with the tobacco industry, but rather the important methodological criticisms posited by three leading statistical scientists, Joseph Berkson, Jerzy Neyman, and Sir Ronald Fisher. Hill and Doll's success in showing that tobacco smoking causes lung cancer required sufficient rebuttal to these critics. Hill's 1965 speech is often cited for its articulation of nine factors to consider in evaluating an association, but the necessary condition is often overlooked. In his speech, Hill identified the situation before the nine factors come into play:

> "Disregarding then any such problem in semantics we have this situation. Our observations reveal an association between two variables, perfectly clear-cut and beyond what we would care to attribute to the play of chance. What aspects of that association should we especially consider before deciding that the most likely interpretation of it is causation?"[124]

The starting point, before the Bradford Hill nine factors come into play, requires a "clear-cut" association, which is "beyond what we would care to attribute to the play of chance." What is "clear-cut" association? The most reasonable interpretation of Bradford Hill is that the starting point is an association that is not the result of chance, bias, or confounding.

Looking at the state of the science after the HSP was published, there were two studies that failed to find any association between PPA and HS. The HSP authors "suggested" an association between PPA appetite suppressant and HS, but with six cases and one control, this was hardly beyond the play of chance. And none of the putative associations were "clear cut" in removing bias and confounding as an explanation for the observations.

H. And Then Litigation Cometh

A tsunami of state and federal cases followed the publication of the HSP study.[125] The Judicial Panel on Multi-district Litigation gave Judge Barbara Rothstein, in the Western District of Washington, responsibility for the pre-trial management of the federal PPA cases. Given the problems with the HSP, the defense unsurprisingly lodged Rule 702 challenges to plaintiffs' expert witnesses' opinions.[126]

---

[124] Austin Bradford Hill, "The Environment and Disease: Association or Causation?" 58 *Proc. Royal Soc'y Med*. 295, 295 (1965).

[125] *See* Barbara J. Rothstein, Francis E. McGovern, and Sarah Jael Dion, "A Model Mass Tort: The PPA Experience," 54 *Drake L. Rev*. 621 (2006); Linda A. Ash, Mary Ross Terry, and Daniel E. Clark, *Matthew Bender Drug Product Liability* § 15.86 PPA (2003).

[126] *In re Phenylpropanolamine Prods. Liab. Litig*., 289 F.Supp. 2d 1230 (W.D. Wash. 2003). Curiously, the defense did not appear to challenge reliance upon the HSP, under Rule 703.

In June 2003, Judge Rothstein issued her decision on the defense motions. After reviewing a selective regulatory history of PPA, the court turned to epidemiology, and its statistical analysis. Although misunderstanding of p-values and confidence intervals is endemic among the judiciary, the descriptions provided by Judge Rothstein portended a poor outcome:

> "P-values measure the probability that the reported association was due to chance, while confidence intervals indicate the range of values within which the true odds ratio is likely to fall."[127]

Both descriptions are seriously incorrect,[128] which is especially concerning given that Judge Rothstein would go on, in 2003, to become the director of the Federal Judicial Center, where she would oversee work on third edition of the *Reference Manual on Scientific Evidence*.

The MDL court also managed to make a mash out of the one-tailed test used in the HSP report. That report was designed to inform regulatory action, where actual conclusions of causation are not necessary. When the HSP authors submitted their paper to the *New England Journal of Medicine*, they of course had to comply with the standards of that journal, and they doubled their reported p-values to comply with the journal's requirement of using a two-tailed test. Some key results of the HSP no longer had p-values below 5 percent, as the defense was keen to point out in its briefings.

From the sources it cited, the court clearly did not understand the issue, which was the need to control for random error. The court declared that it had found:

> "that the HSP's one-tailed statistical analysis complies with proper scientific methodology, and concludes that the difference in the expression of the HSP's findings [and in the published article] falls far short of impugning the study's reliability."[129]

This finding ignores the very different contexts between regulatory action and causation in civil litigation. The court's citation to the second edition of the *Reference Manual on Scientific Evidence* further illustrates its confusion:

> "Since most investigators of toxic substances are only interested in whether the agent increases the incidence of disease (as distinguished from providing protection from the disease), a one-tailed test is often viewed as appropriate."
> *****

---

[127] Id. at 1236 n.1

[128] Michael O. Finkelstein & Bruce Levin, *Statistics for Lawyers* 171, 173-74 (3rd ed. 2015). *See also* Sander Greenland, Stephen J. Senn, Kenneth J. Rothman, John B. Carlin, Charles Poole, Steven N. Goodman, and Douglas G. Altman, "Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations," 31 *Eur. J. Epidem.* 337 (2016).

[129] *In re Phenylpropanolamine Prods. Liab. Litig.*, 289 F.Supp. 2d 1230, 1241 (W.D. Wash. 2003).

> "a rigid rule [requiring a two-tailed test] is not required if p-values and significance levels are used as clues rather than as mechanical rules for statistical proof."[130]

In a sense, given the prevalence of advocacy epidemiology, many researchers are interested in only showing an increased risk. Nonetheless, the point of evaluating p-values is to assess random error involved in sampling of a population, and that sampling generates a rate of error even when the null hypothesis is assumed to be absolutely correct. Random error can go in either direction, resulting in risk ratios above or below 1.0. Indeed, the probability of observing a risk ratio of exactly 1.0, in a large study, is incredibly small even if the null hypothesis is correct. The risk ratio for men who had used a PPA product was below 1.0, which also recommends a two-tailed test. Trading on the confusion of regulatory and litigation findings, the court proceeded to mischaracterize the parties' interests in designing the HSP, as only whether PPA increased the risk of stroke. In the PPA MDL, the parties did not want "clues," or help on what FDA policy should be; they wanted a test of the causal hypothesis.

In a footnote, the court pointed to testimony of Dr. Ralph Horwitz, one of the HSP investigators, who stated that "[a]ll parties involved in designing the HSP were interested solely in testing whether PPA increased the risk of stroke." The parties, of course, were not designing the HSP for support for litigation claims.[131] The court also cited, in this footnote, a then recent case that found a one-tailed p-value inappropriate "where that analysis assumed the very fact in dispute." The plaintiffs' reliance upon the one-sided p-values in the unpublished HSP report did exactly that.[132] The court tried to excuse the failure to rule out random error by pointing to language in the published HSP article, where the authors stated that inconclusive findings raised "concern regarding safety."[133]

In analyzing the defense challenge to the opinions based upon the HSP, Judge Rothstein committed both legal and logical fallacies. First, citing Professor David Faigman's treatise for the proposition that epidemiology is widely accepted because the "general techniques are valid," the court found that the HSP, and reliance upon it, was valid despite the identified problems. The issue, however, was not whether epidemiological techniques are valid, but whether the techniques used in the HSP were valid. The devilish details of the HSP in particular largely went ignored.[134] From a legal perspective, Judge Rothstein's opinion can be seen to place a burden

---

[130] *Id*. (citing *Reference Manual* at 126-27, 358 n. 69). The edition of *Manual* was not identified by the court.
[131] *Id*. at n.9, citing deposition of Ralph Horowitz [sic].
[132] *Id*., citing *Good v. Fluor Daniel Corp*., 222 F.Supp. 2d 1236, 1242-43 (E.D. Wash. 2002).
[133] *Id*. 1241, citing Kernan, *supra* note 95, at 183.
[134] *In re Phenylpropanolamine Prods. Liab. Litig*., 289 F.Supp. 2d 1230, 1239 (W.D. Wash. 2003) (citing 2 *Modern Scientific Evidence: The Law and Science of Expert Testimony* § 28-1.1, at 302-03 (David L. Faigman, *et al*., eds., 1997) ("Epidemiologic studies have been well received by courts trying mass tort suits. Well-conducted studies are uniformly admitted. The widespread acceptance of epidemiology is based in large part on the belief that the general techniques are valid.").

upon the defense to show invalidity, by invoking a presumption of validity. This shifting of the burden was then, and is now, contrary to the law.

Perhaps the most obvious dodge of the court's gatekeeping responsibility came with the conclusory assertion that the "Defendants' *ex post facto* dissection of the HSP fails to undermine its reliability. Scientific studies almost invariably contain flaws."[135] Perhaps it is sobering to consider that all human beings have flaws, and yet somehow we distinguish between sinners and saints, and between criminals and heroes. The court shirked its responsibility to look at the identified flaws to determine whether they threatened the HSP's internal validity, as well as its external validity in the plaintiffs' claims for hemorrhagic strokes in each of the many subgroups considered in the HSP, as well as outcomes not considered, such as myocardial infarction and ischemic stroke. Given that there was but one key epidemiologic study relied upon for support of the plaintiffs' extravagant causal claims, the identified flaws might have been expected to lead to some epistemic humility.

The PPA MDL court exhibited a willingness to cherry pick HSP results to support its low-grade gatekeeping. For instance, the court recited that "[b]ecause no men reported use of appetite suppressants and only two reported first use of a PPA-containing product, the investigators could not determine whether PPA posed an increased risk for hemorrhagic stroke in men."[136] There was, of course, another definition of PPA exposure that yielded a total of 19 exposed men, about one-third of all exposed cases and controls. For the men with any exposure to OTC PPA cough cold remedies, there were six male cases with HS, and 13 controls, with a reported odds ratio of 0.62 (95%, C.I., 0.20 – 1.92); p = 0.41. Although the result for men was not statistically significant, and the interval is wide, the point estimate for the sample was a risk ratio below one, with a confidence interval that excludes a doubling of the risk based upon this sample statistic. The number of male HS exposed cases was the same as the number of female HS appetite suppressant cases, which somehow did not disturb the court.

Superficially, the PPA MDL court appeared to place great weight on the fact of peer review publication in a prestigious journal, by well-credentialed scientists and clinicians. Given that "[t]he prestigious NEJM published the HSP results … research bears the indicia of good science."[137] This judgment differs remarkably from that of the *New England Journal of Medicine* editor Marcia Angell, who observed that "peer review is not and cannot be an objective scientific

---

[135] *Id*. at 1240. The court cited the *Reference Manual on Scientific Evidence* 337 (2d ed. 2000), for this universal attribution of flaws to epidemiology studies ("It is important to recognize that most studies have flaws. Some flaws are inevitable given the limits of technology and resources.") Of course, when technology and resources are limited, expert witnesses are permitted to say "I cannot say." The PPA MDL court also cited another MDL court, which declared that "there is no such thing as a perfect epidemiological study." *In re Orthopedic Bone Screw Prods. Liab. Litig*., MDL No. 1014, 1997 WL 230818, at *8-9 (E.D.Pa. May 5, 1997).

[136] *Id*. at 1236.

[137] *Id*. at 1239.

process, nor can it be relied on to guarantee the validity or honesty of scientific research, despite much uninformed opinion to the contrary."[138]

Although Professor Susan Haack's writings on law and science are idiosyncratic, her snarky analysis of this kind of blind reliance on peer review is noteworthy:

> "though peer-reviewed publication is now standard practice at scientific and medical journals, I doubt that many working scientists imagine that the fact that a work has been accepted for publication after peer review is any guarantee that it is good stuff, or that it's not having been published necessarily undermines its value. The legal system, however, has come to invest considerable epistemic confidence in peer-reviewed publication — perhaps for no better reason than that the law reviews are not peer-reviewed!"[139]

Ultimately, the PPA MDL court revealed that it was quite inattentive to the validity concerns of the HSP. Among the cases filed in the federal court were heart attack and ischemic stroke claims. The HSP did not address those claims, and the MDL court was perfectly willing to green light the claims on the basis of case reports and expert witness hand waving about "plausibility." Not only was this reliance upon case reports plus biological plausibility against the weight of legal authority, it was against the weight of scientific opinion, as expressed by the HSP authors themselves:

> "Although the case reports called attention to a possible association between the use of phenylpropanolamine and the risk of hemorrhagic stroke, the absence of control subjects meant that these studies could not produce evidence that meets the usual criteria for valid scientific inference"[140]

Since no epidemiology was necessary at all for ischemic stroke and myocardial infarction claims, then a deeply flawed epidemiologic study was thus even better than nothing for any HS claim. The court's strained reasoning revealed that peer review and prestige were merely window dressing.

The HSP study was subjected to much greater analysis in actual trial litigation. Before the MDL court concluded its abridged gatekeeping, the defense successfully sought the underlying data to the HSP. Plaintiffs' counsel and the Yale investigators resisted and filed motions to quash the

---

[138] Arnold S. Relman & Marcia Angell, "How Good is Peer Review?"321 *New Engl. J. Med*. 827, 828 (1989).

[139] Susan Haack, "Irreconcilable Differences? The Troubled Marriage of Science and Law," 72 *Law & Contemp. Problems* 1, 19 (2009) (internal citations omitted). It may be telling that Haack has come to publish much of her analysis in law reviews. *See* Nathan Schachtman, "Misplaced Reliance On Peer Review to Separate Valid Science From Nonsense" *Tortini* (Aug. 14, 2011).

[140] Kernan, *supra* note 95, at 1831.

defense subpoenas. The MDL court denied the motions and required the parties to collaborate on redaction of medical records to be produced.[141]

In a law review article published a few years after the PPA Rule 702 decision, Judge Rothstein immodestly described the PPA MDL as a "model mass tort," and without irony characterized herself as having taken "an aggressive role in determining the admissibility of scientific evidence [].”[142]

The MDL PPA Rule 702 decision stands as a landmark of judicial incuriousness and credulity. The court conducted hearings and entertained extensive briefings on the reliability of plaintiffs' expert witnesses' opinions, which were based largely upon one epidemiologic study, HSP. In the end, publication in a prestigious peer-reviewed journal was used as a complete substitute for independent review and critical analysis.[143] The admissibility challenges were refused.

I. Exuberant Praise for Judge Rothstein

In 2009, an American Law Institute – American Bar Association continuing legal education seminar on expert witnesses and environmental litigation, Anthony Roisman presented on "*Daubert* & Its Progeny - Finding & Selecting Experts - Direct & Cross-Examination." Roisman has been active in various plaintiff advocacy organizations, including serving as the head of the American Trial Lawyers' Association Section on Toxic, Environmental & Pharmaceutical Torts (STEP). In his 2009 lecture, Roisman praised Judge Rothstein's PPA Rule 702 decision as "the way *Daubert* should be interpreted." More concerning was Roisman's revelation that Judge Rothstein wrote the PPA decision, "fresh from a seminar conducted by the Tellus Institute, which is an organization set up of scientists to try to bring some common sense to the courts' interpretation of science, which is what is going on in a *Daubert* case."[144]

---

[141] *In re Propanolamine Prods. Litig*., MDL 1407, Order re Motion to Quash Subpoenas re Yale Study's Hospital Records (W.D. Wash. Aug. 16, 2002). Two of the HSP investigators wrote an article, over a decade later, to complain about litigation efforts to obtain data from ongoing studies. They did not mention the PPA case. Walter N. Kernan, Catherine M. Viscoli, and Mathew C. Varughese, "Litigation Seeking Access to Data From Ongoing Clinical Trials: A Threat to Clinical Research," 174 *J. Am. Med. Ass'n Intern. Med*. 1502 (2014).

[142] Barbara J. Rothstein, Francis E. McGovern, and Sarah Jael Dion, "A Model Mass Tort: The PPA Experience," 54 *Drake L. Rev*. 621, 638 (2006).

[143] *In re Phenylpropanolamine Prods. Liab. Litig*., 289 F.Supp. 2d 1230, 1239 (W.D. Wash. 2003) (proposing that peer review shows that the challenged research meets the minimal criteria for good science).

[144] Anthony Roisman, "*Daubert* & Its Progeny - Finding & Selecting Experts - Direct & Cross-Examination," ALI-ABA 2009. Roisman's remarks about the role of Tellus Institute start just after minute 8, on the recording, available from the American Law Institute, and the author.

Roisman's endorsement of the PPA decision may have been nothing more than emotive, result-oriented approval, but what of his enthusiasm for the "learning" that Judge Rothstein received fresh from the Tellus Institute? What exactly is or was the Tellus Institute?

In June 2003, the same month as Judge Rothstein's PPA decision, the Tellus Institute supported a group known as Scientific Knowledge and Public Policy (SKAPP) in publishing an attack on the *Daubert* decision. The Tellus-SKAPP paper, "*Daubert*: The Most Influential Supreme Court Ruling You've Never Heard Of," appeared online in 2003.[145]

David Michaels, a plaintiffs' expert in chemical exposure cases, and a founder of SKAPP, has typically described his organization as having been funded by the Common Benefit Trust, "a fund established pursuant to a court order in the Silicone Gel Breast Implant Liability litigation."[146] What Michaels hides is that this "Trust" is nothing other than the common benefits fund set up in MDL 926, as it is for most MDLs, to permit plaintiffs' counsel to retain and present expert witnesses in the common proceedings. In other words, it was the plaintiffs' lawyers' walking-around money. SKAPP's sister organization, the Tellus Institute, was clearly aligned with SKAPP. Alas, Richard Clapp, who was a testifying expert witness for PPA plaintiffs, was an active member of the Tellus Institute at the time of the judicial educational seminar for Judge Rothstein.[147] Clapp is listed as a member of the planning committee responsible for preparing the anti-*Daubert* pamphlet. In 2005, as director of the Federal Judicial Center, Judge Rothstein attended another conference, "the Coronado Conference, which was sponsored by SKAPP.[148]

Roisman's revelation in 2009, after the dust had settled on the PPA litigation, may well put Judge Rothstein in the same category as Judge James Kelly, against whom the U.S. Court of Appeals for the Third Circuit issued a writ of mandamus for recusal. Judge Kelly was invited to attend a conference on asbestos medical issues, set up by Dr. Irving Selikoff with scientists who testified for plaintiffs' counsel. The conference was funded by plaintiffs' counsel, Ron Motley.

Roisman was counsel of record for the losing side in *Kumho Tire Co. v. Carmichael*, 526 U.S. 137 (1999).

[145] *See* "*Daubert*: The Most Influential Supreme Court Ruling You've Never Heard Of; A Publication of the Project on Scientific Knowledge and Public Policy, coordinated by the Tellus Institute" (2003).

[146] *See, e.g.*, David Michaels, *Doubt is Their Product: How Industry's War on Science Threatens Your Health* 267 (2008).

[147] *See* Richard W. Clapp & David Ozonoff, "Environment and Health: Vital Intersection or Contested Territory?" 30 *Am. J. L. & Med*. 189, 189 (2004) ("This Article also benefited from discussions with colleagues in the project on Scientific Knowledge and Public Policy at Tellus Institute, in Boston, Massachusetts.").

[148] *See* Barbara Rothstein, "Bringing Science to Law," 95 Am. J. Pub. Health S1 (2005) ("The Coronado Conference brought scientists and judges together to consider these and other tensions that arise when science is introduced in courts.").

The co-conspirators, Selikoff and plaintiffs' counsel, paid for Judge Kelly's transportation and lodgings, without revealing the source of the funding.[149]

In the case of Selikoff and Motley's effort to subvert the neutrality of Judge James M. Kelly in the school district asbestos litigation, and pervert the course of justice, the conspiracy was detected in time for a successful recusal effort. In the PPA litigation, there was no disclosure of the efforts by the anti-*Daubert* advocacy group, the Tellus Institute, to undermine the neutrality of a federal judge.

J. Aftermath of Failed MDL Gatekeeping

Ultimately, the HSP study received much more careful analysis before juries. Although the cases that went to trial involved plaintiffs with catastrophic injuries and a high-profile article in the *New England Journal of Medicine*, the jury verdicts were overwhelmingly in favor of the defense.[150]

In the first case that went to trial (but second to verdict), the defense presented a thorough scientific critique of the HSP. The underlying data and medical records that had been produced in response to a Rule 45 subpoena in the MDL allowed juries to see that the study investigators had deviated from the protocol in ways to increase the number of exposed cases, with the obvious result of increasing the odds ratios reported. Juries were ultimately much more curious about evidence and testimony on reclassifications of exposure that drove up the odds ratios for PPA use, than they were about the performance of linear logistic regressions.

The HSP investigators were well aware of the potential for medication use to occur after the onset of stroke symptoms (headache), which may have sent a person to the medicine chest for an OTC cold remedy. At least three of the female "first use" cases involved exposure differential misclassification. Case 71-0039 was just such a case, as shown by the medical records and the HSP investigators' initial classification of the case. On dubious grounds, however, the HSP investigators reclassified stroke onset to after PPA-medication use, in what the investigators knew increased their chances of finding an association.

The reclassification of Case 20-0092 was even more egregious. The patient was originally diagnosed as having experienced a transient ischemic attack (TIA), after a CT of the head showed no bleed. Case 20-0092 was not even a case. The patient's TIA was treated with heparin,

---

[149] *In re School Asbestos Litigation*, 977 F.2d 764 (3d Cir. 1992). *See* Cathleen M. Devlin, "Disqualification of Federal Judges – Third Circuit Orders District Judge James McGirr Kelly to Disqualify Himself So As To Preserve 'The Appearance of Justice' Under 28 U.S.C. § 455 – In re School Asbestos Litigation (1992)," 38 *Villanova L. Rev*. 1219 (1993); Bruce A. Green, "May Judges Attend Privately Funded Educational Programs? Should Judicial Education Be Privatized?: Questions of Judicial Ethics and Policy," 29 *Fordham Urb. L.J*. 941, 996-98 (2002).

[150] Alison Frankel, "A Line in the Sand," *The Am. Lawyer - Litigation* (2005); Alison Frankel, "The Mass Tort Bonanza That Wasn't," *The Am. Lawyer* (Jan. 6, 2006).

an appropriate therapy for ischemic stroke, but one that is known to cause bleeding. The following day, MRI of the head revealed a HS. The HSP classified Case 20-0092 as a case.

In Case 18-0025, the patient experienced a headache in the morning, and took a PPA-medication (Contac) for relief. The stroke was already underway when the Contac was taken, but the HSP reversed the order of events.

Case 62-0094 presented an interesting medical history that included an event no one in the HSP considered including in the interview protocol. In addition to a history of heavy smoking, alcohol, cocaine, heroin, and marijuana use, and a history of seizure disorder, Case 62-0094 suffered a traumatic head injury immediately before developing a SAH. Treating physicians ascribed the SAH to traumatic injury, but understandably there were no controls that were identified with similar head injury within the exposure period.

Plaintiffs' expert witness Richard Clapp accused the defense of "hacking at the A cell," but juries seemed to understand that the hacking had started before the paper was published. The facts of the exposed HS cases were presented in detail in a trial of two consolidated cases, in Los Angeles County. After the jury returned a verdict for the defense in both of the plaintiffs' cases, plaintiffs' counsel challenged the defendant's reliance upon underlying data in the HSP, which went behind the peer-reviewed publication, and which showed that the peer review failed to prevent serious errors. In essence, the plaintiffs' counsel claimed that the defense experts' scrutiny of the underlying data and investigator misclassifications was itself not "generally accepted" methodology, and thus inadmissible under California law. The trial court rejected the plaintiffs' claim and their request for a new trial, and spoke to the significance of challenging the superficial and ineffective peer review of the key study relied upon by plaintiffs in the PPA litigation:

> "I mean, you could almost say that there was some unethical activity with that Yale Study. It's real close. I mean, I — I am very, very concerned at the integrity of those researchers.
>
> ********
>
> Yale gets — Yale gets a big black eye on this."[151]

Epidemiologist Charles Hennekens, who had been a consultant to the PPA-medication manufacturers, published a critique of the HSP study, in 2006. The Hennekens' critique included many of the criticisms lodged by himself, as well as by epidemiologists Lewis Kuller, Noel Weiss, and Brian Strom, back in an October 2000 FDA meeting, before the HSP was published. Richard Clapp, Tellus Institute activist and expert witness for PPA plaintiffs, and Michael

[151] *O'Neill v. Novartis AG*, California Superior Court, Los Angeles Cty., Transcript of Oral Argument on Post-Trial Motions, at 46 -47 (March 18, 2004) (Hon. Anthony J. Mohr), *aff'd sub nom. O'Neill v. Novartis Consumer Health, Inc.*,147 Cal. App. 4th 1388, 55 Cal. Rptr. 3d 551, 558-61 (2007).

Williams, lawyer for PPA claimants, wrote a letter criticizing Hennekens.[152] David Michaels, an expert witness for plaintiffs in other chemical exposure cases, and a founder of SKAPP, which collaborated with the Tellus Institute on its anti-*Daubert* compaign, wrote a letter accusing Hennekens of "mercenary epidemiology," for engaging in re-analysis of a published study. Michaels never complained about the litigation-inspired re-analyses put forward by plaintiffs' witnesses in the Bendectin litigation. Plaintiffs' lawyers and their expert witnesses had much to gain by starting the litigation and trying to expand its reach. Defense lawyers and their expert witnesses effectively put themselves out of business by shutting it down.[153]

Neither the Yale HSP nor the PPA MDL court decision was ever retracted.

VII. Conclusion

The silicone and PPA litigations are hardly unique in involving studies with inaccurate, misleading data and methods. Similar experiences can be recounted from litigation of claims of isotretinoin and suicide,[154] acetaminophen and liver failure,[155] welding and parkinsonism,[156] sildenafil and ophthalmic events,[157] and MMR vaccine and autism,[158] to name a few. The Russian proverb, доверяй, но проверяй, "trust but verify," is wise counsel in both the world of science and of litigation.

---

[152] Richard Clapp & Michael L. Williams, Regarding ''Phenylpropanolamine and Hemorrhagic Stroke in the Hemorrhagic Stroke Project,'' 16 *Ann. Epidem*. 580 (2006).

[153] David Michaels, "Regarding 'Phenylpropanolamine and Hemorrhagic Stroke in the Hemorrhagic Stroke Project': Mercenary Epidemiology - Data Reanalysis and Reinterpretation for Sponsors with Financial Interest in the Outcome," 16 *Ann. Epidem*. 583 (2006). Hennekens responded to these letters. Stier & Hennekens, note 109, *supra*.

[154] *Palazzolo v. Hoffman La Roche, Inc*., 2010 WL 363834, *5 (N.J. App. Div. 2010). Discovery revealed that the study author, James Bremner, did not follow the methodology described in his paper, and that he could not document the data used in the paper's analysis, or the correctness of his statistical analyses. The New Jersey Appellate Division held that Bremner's study was not sound and reliably generated, which precluded reliance upon it.

[155] *In re Tylenol (Acetaminophen) Marketing, Sales Practices and Prods. Liab. Litig*., MDL No. 2436 (E.D.Pa.).

[156] *In re Welding Rod Prods. Liab. Litig*., No. 1:03–CV–17000, MDL No. 1535, 2005 WL 5417815 (Oct. 18, 2005). After the defense had the opportunity to analyze a limited set of data produced by Dr. Brad Racette, plaintiffs' expert witnesses abjured reliance upon his study.

[157] *In re Viagra Prods. Liab. Litig*., 658 F. Supp. 2d 936, 945 (D. Minn. 2009) ("[p]eer review and publication mean little if a study is not based on accurate underlying data'').

[158] *See* Andrew J. Wakefield, *et al*., "Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children," 351 *Lancet* 637 (1998); Editors of the Lancet, "Retraction—Ileal-lymphoidnodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children," 375 *Lancet* 445 (2010).

The last half century has seen both science and law on a trajectory that requires greater attention to validity of data collection, analysis, and inference, and consequently to the threats to validity from questionable research practices. There are, of course, strong economic incentives to dilute this attention. In the world of science, published articles are the tokens of productivity and success in the academy and before public and private granting institutions. In law, expert witnesses build opinions from published articles and by relying upon them, vouch for their validity. Outside a few unenlightened state courtrooms, the quality of the published literature has become a close concern for lawyers, regulators, and policy makers. The prevalence of QRPs, expressions of concern, and retractions has become everyone's business.