

The Curious Case of *Wendell v. GlaxoSmithKline LLC*

David L. Faigman & Jennifer Mnookin*

I. INTRODUCTION	607
II. CURIUSER AND CURIUSER	609
III. EVALUATING SCIENTIFIC EVIDENCE	612
A. Scientific Evidence That is Based on Art, Not Science..	612
B. Relying on Qualifications Alone	615
C. What is the “Task at Hand” in Medical Causation Testimony?	617
IV. CONCLUSION	627

I. INTRODUCTION

All evidence introduced in court must meet some threshold standard in order to be admitted. Even the lowliest of proof must pass a relevancy test. In the context of expert evidence under Federal Rule of Evidence 702, the precise threshold that must be met remains in considerable doubt, even nearly twenty-five years after the Supreme Court sought to clarify the standard in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*¹ The single-most important principle announced in *Daubert* was that some measure of scientific realism should guide trial court admissibility decisions. In the case of scientific evidence, for instance, “scientific knowledge” and “scientific validity” were the prescribed guidelines.² But in respect to all expert evidence, trial courts were required to evaluate the underlying bases for the proffered expert opinion to assess whether it was adequately valid and

* David Faigman, Chancellor & Dean and John F. Digardi Distinguished Professor of Law at the University of California, Hastings College of the Law and Professor at the University of California, San Francisco, School of Medicine. Jennifer Mnookin, Dean and David G. Price and Dallas P. Price Professor of Law at the UCLA School of Law. Many thanks to Jodi Kruger and her fellow UCLA librarians for very helpful support and assistance.

¹ *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579 (1993). *Daubert*, decided in 1993, is often used as shorthand for the current Rule 702, which was amended in 2000 largely to codify the standard set forth in that case. Hence, the applicable rule is 702, but the *Daubert* decision continues to be cited for having set minimum expectations for the quality of expert opinion evidence offered in court.

² *Id.* at 589–92.

reliable.³ For all expert opinions, the trial court was to be the gatekeeper to ensure that there were “good grounds” for an expert’s proffered evidence.⁴

But this threshold gatekeeping requirement, though now well-ensconced, continues to raise as many questions as it answers. How high should the threshold be? Just how much scientific support must ground an expert’s opinion to warrant admissibility? How should a court assess whether experts know what they think they know? And, moreover, how should courts go about determining whether the expert’s knowledge is based on sufficiently good grounds to be admitted? These questions have particular salience when experts wish to testify to scientifically plausible but weakly supported claims. Furthermore, in many product liability and toxic torts cases, admissibility decisions about the expert’s testimony may well be case-dispositive: admit the expert evidence, and the case—often involving a tragic set of facts—comes before the jury, but exclude that evidence, and summary judgment is a foregone conclusion because no admissible evidence supports causation. When causation is possible but scientifically unproven, how should the courts respond? If courts take seriously *Daubert*’s holding that courts should employ scientific sensibilities to evaluate expert opinion evidence, exclusion and summary judgment are doctrinally required when adequate scientific support does not establish causation. But when tantalizing glimmers of evidence suggest that causation is possible, albeit far from scientifically established, some courts respond by resisting, or stretching, *Daubert*’s strictures. In this Article, we consider the curious case of *Wendell v. GlaxoSmithKline*, a Ninth Circuit decision that fails so dramatically to employ scientific reasoning that it serves well as a cautionary tale.

Wendell also illustrates how, when courts are making admissibility decision under conditions of scientific uncertainty, their reasoning matters as much as the conclusion. To put the point starkly: the Ninth Circuit’s reasoning in *Wendell* is indefensible, but its conclusion *might* not be – at least if it had been making an admissibility judgment in the first instance, rather than on appeal. On appeal, however, we see little justification for overturning the trial court’s judgment, given the paucity of scientific evidence to support causation and the appropriate standard of review – and certainly, the Ninth Circuit’s reasoning on this score is not persuasive.

When we began writing this Article, the Supreme Court had yet to decide whether to review the Ninth Circuit’s holding. Ultimately, the Court did not grant cert.⁵ Nonetheless, we expect that the Court will need to engage

³ *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 149 (1999).

⁴ *Daubert*, 509 U.S. at 589–90.

⁵ The case, renamed *Teva Pharmaceuticals, Inc. v. Wendell*, has been scheduled for conference for March 16, 2018. See SCOTUSBLOG (Mar. 19, 2018), <http://www.scotusblog.com>

further with questions about the application of Rule 702 and Daubert to the many thorny issues of causation in toxic tort cases. In what follows, we therefore use *Wendell* as a case study and an object lesson, but the important issues raised by this case about how to – and how not to – assess causation in toxic torts cases have implications well beyond this particular instance and example.

II. CURIUSER AND CURIUSER

Like many such cases, the underlying story of *Wendell* is heartbreakingly tragic. Maxx Wendell died at the age of twenty-one of Hepatosplenic T-cell lymphoma (HSTCL), a very rare and aggressive cancer. Prior to developing HSTCL, Maxx was treated with a variety of drugs for inflammatory bowel disease (IBD), including in particular, mercaptopurine (6-MP). Maxx's parents sued the manufacturers of these drugs, claiming that 6-MP interacted with another drug, the tumor necrosis factor alpha antagonist (anti-TNF), to cause Maxx's HSTCL.

Two highly qualified physicians “opined in their expert reports that the combination of 6-MP drugs and anti-TNF drugs prescribed to Maxx increased his likelihood of developing HSTCL and, ultimately, caused his death.”⁶ But the district court found that these opinions “are not based on sufficiently reliable scientific data,” and hence, could not be admissible under Rule of Evidence 702 and *Daubert*. The district court granted the defendant's motion for summary judgment, finding that the plaintiffs had “failed to present sufficient evidence to support an inference that [6-MP], either alone or in combination with anti-TNF drugs, caused Maxx to develop HSTCL.”⁷ There was no dispute about whether the experts were qualified or whether their opinions were relevant. But the district court appropriately recognized that the key evidentiary question was neither of these, but rather, whether the experts' opinions themselves were adequately reliable.

The district court had little difficulty answering that question in the negative. As the judge explained, plaintiffs' experts conceded that there were neither animal studies nor any epidemiological studies indicating the relationship between 6-MP and anti-TNF drugs and the development of HSTCL. Moreover, the experts acknowledged that they did not employ the same level of rigor in forming their expert testimony that they would use for reaching similar determinations directed at other scientists rather than the court, through publication in peer-reviewed journals rather than testimony.

om/case-files/cases/teva-pharmaceuticals-usa-inc-v-wendell/.

⁶ *Wendell v. Johnson & Johnson*, No. 09-cv-04124, 2014 WL 2943572, at *3 (N.D. Cal. June 30, 2014), *rev'd sub nom.* *Wendell v. GlaxoSmithKline LLC*, 858 F.3d 1227 (9th Cir. 2017).

⁷ *Id.* at *7.

For instance, when one expert “was asked whether his opinions in this case would be publishable in a medical article, he replied that the standard for publication would ‘probably be more rigorous’ than the standard he applied in forming his opinions.”⁸ In addition, more than seventy percent of HSTCL cases, according to the experts, are idiopathic, meaning that the cause is unknown. This, the trial judge correctly realized, makes it impossible to determine cause reliably simply by making use of a differential etiology method, whereby other known causes are eliminated;⁹ if the main cause is “unknown,” eliminating other known causes establishes little, absent “some reliable evidence of a positive link between the drugs at issue and the disease.”¹⁰ To be sure, some case reports and incident studies the experts relied upon described other examples of HSTCL—acknowledged to be an extremely rare cancer—among other patients who also had taken a combination of 6-MP and anti-TNF drugs (though not necessarily the precise same drugs as Maxx). But for the trial court, these reports amounted to suggestive anecdotes, not scientific proof—nor, according to the court, had the physicians eliminated IBD itself as a potential cause of Maxx’s cancer.

Up to this point, the story is familiar. The plaintiffs’ experts relied on a generalized clinical judgment, informed largely by anecdote and conjecture, perhaps plausible, but insufficiently proven. While the experts were willing to testify to a “reasonable degree of medical certainty” about cause, even they acknowledged their causal conjecture didn’t rise to the standards of peer review in a scientific journal. Since the law requires that the underlying basis for the experts’ opinion be adequately supported by “scientific knowledge,” the trial court excluded their testimony. Once their experts’ opinions were excluded, the plaintiffs lacked sufficient—or indeed any—admissible evidence of causation, so the defendant’s summary judgment motion was, naturally, granted. Additionally, given the holding in *General Electric Company v. Joiner*,¹¹ which mandates an abuse of discretion standard for the appellate review of trial courts’ admissibility decisions involving expert evidence, the appellate outcome appeared well-determined. In light of *Daubert* and Rule 702’s insistence that scientific opinions in court require an adequate scientific basis, how could a trial judge who excludes expert opinions based solely on clinical judgment, case reports, and conjecture (and without animal studies, epidemiological studies, or a persuasive differential etiology) be seen to have abused his discretion?

⁸ *Id.* at *4.

⁹ See *infra* notes 22–23 for a more complete discussion of the methods associated with “differential etiology.”

¹⁰ *Wendell*, 2014 WL 2943572, at *5.

¹¹ *Gen. Elec. Co. v. Joiner*, 522 U.S. 136 (1997).

At the Ninth Circuit, however, is where the case gets “curiouser and curiouser.”¹² The appellate court at least nominally accepted that its review was limited to the highly deferential abuse of discretion standard.¹³ But the court went on to warn that it reviews *de novo* the “construction or interpretation of . . . the Federal Rules of Evidence, including whether particular evidence falls within the scope of a given rule.”¹⁴

Unfortunately, the appellate court never returned to the question of what standard of review it employed to overturn the trial court’s decision, nor precisely what it meant by its statement that it had the power to review *de novo* “whether particular evidence falls within the scope of a given rule.” On its face, of course, the testimony of the plaintiffs’ two experts unambiguously fell within the scope of Rule 702. Thus, the lower court’s finding that the evidence here—i.e., expert medical causation evidence—was subject to Rule 702 was plainly not error. While the appropriate scope of Rule 702 could indeed be a legal question warranting *de novo* review, that hardly seems applicable to these facts; there was no plausible argument that Rule 702 did not apply to the causation evidence proffered by the plaintiffs’ experts. A more ambitious reading of the appellate court’s assertion—that is, that all admissibility decisions regarding whether expert evidence is admissible “within the scope of the rule” is subject to *de novo* review—would effectively nullify the *Joiner* ruling, and the court did not seem to mean to go that far with its oblique sentence gesturing to some possible, not fully specified, role for a *de novo* standard.

In any event, while the Ninth Circuit’s conclusions largely used the rhetorical register of “abuse of discretion,” both its reasoning and its conclusions are hard to justify under that deferential standard. The court concludes that while it is a “close question,” the district court “erred” and “abused its discretion by excluding Dr. Shustov’s and Dr. Weisenburger’s testimony. . . .”¹⁵ Even if formally engaged in an abuse of discretion analysis, the court did not seem to show the kinds of deference to the district court that the standard would typically invite.

Why did the Ninth Circuit reach the opposite conclusion from the trial court? Without question, the plaintiffs’ experts were two highly qualified and well-regarded physicians.¹⁶ This fact turned out to be the primary basis for the appellate court’s belief that the lower court had erred. The court

¹² LEWIS CARROLL, ALICE’S ADVENTURES IN WONDERLAND AND THROUGH THE LOOKING GLASS 16 (Lothrop Publishing Co., 1898) (1865).

¹³ *Wendell v. GlaxoSmithKline LLC*, 858 F.3d 1227, 1231 (9th Cir. 2017).

¹⁴ *Id.* at 1231 (alteration in original) (quoting *Messick v. Novartis Pharm. Corp.*, 747 F.3d 1193, 1196 (9th Cir. 2014)).

¹⁵ *Id.* at 1233, 1237.

¹⁶ *Id.* at 1233.

summarized its view of the proffered expert evidence as follows:

“[M]edicine partakes of art as well as science.” Where, as here, two doctors who stand at or near the top of their field and have extensive clinical experience with the rare disease or class of disease at issue, are prepared to give expert opinions supporting causation, we conclude that *Daubert* poses no bar based on their principles and methodology.¹⁷

This remarkable paragraph deserves to be dissected and examined in at least three parts. The first involves the court’s broad-brush assertion regarding medicine partaking of art as well as science. The second is the outsized role qualifications played in the court’s assessment. And the third concerns the relationship, if any, between extensive clinical experience and determining medical causation.

III. EVALUATING SCIENTIFIC EVIDENCE

A. *Scientific Evidence That is Based on Art, Not Science*

Whenever someone asserts that “[m]edicine partakes of art as well as science,” we can suspect that science, at least in this instance, is likely being tossed to the wind. It is not that the claim is wrong – the practice of medicine is indeed both an art and a science; but a claim of scientific causation ought to be grounded in valid scientific knowledge and based on reasoned explanations that go beyond medicine as an ‘art.’ The court itself noted earlier in its opinion that “[s]cientific evidence is reliable ‘if the principles and methodology used by an expert are grounded in the methods of science.’”¹⁸ The question presented in *Wendell* was straightforward, even if the answer was anything but. Specifically, the key question was whether or not the defendant’s drug was substantial cause of Maxx’s illness. While medical *treatment* may indeed have artistry associated with it, where the art lies in determining whether adequate scientific support establishes that A *causes* B is far less obvious; and the Ninth Circuit, in any case, did not explain it. Furthermore, *Daubert* and its progeny stand for the idea that expert evidence requires more than “ipse dixit” to be admissible—hunches, conjectures, and even a good nose for diagnosis are not enough to warrant admissibility.

¹⁷ *Id.* at 1237 (quoting *Messick*, 747 F.3d at 1198).

¹⁸ *Id.* at 1232 (quoting *Clausen v. M/V New Carissa*, 339 F.3d 1049, 1056 (9th Cir. 2003)).

Of course, there are times when experience—what the Ninth Circuit likely meant by “art”—should be a basis for, and possibly even a sufficient basis for, admitting expert evidence. The difficulty lies in determining when this is so. Courts regularly admit all sorts of experts based on their experience with a subject area, from auto mechanics to real estate appraisers. These disciplines neither hold themselves out to be scientific, nor do their fields regularly employ scientific methods to test their hypotheses. Nonetheless, common sense recognizes that some fields, and some experts, have relevant and reliable knowledge to impart to judicial proceedings based on their long experience with the subject. The classic example of this is the harbor pilot who has maneuvered a particular waterway hundreds of times and is called upon to provide an expert opinion, say, about the dangers of a sandbar to local shipping. In such cases, experience will indeed provide useful expertise that ordinarily will meet the strictures of Rule 702.¹⁹

At the same time, the courts have a long and troubled history of permitting experienced-based expert testimony in cases that are ostensibly based on scientific techniques, but which have not been validated beyond the everyday experience of the so-called experts. Most notoriously, this has occurred, and continues to occur, in courts’ reception of forensic science.²⁰ Historically, many forensic specialties were admitted based on experience but, when later tested by standard methods of science, turned out to be deeply flawed, often overstated, and sometimes wholly invalid. Some areas no longer admitted in court include microscopic hair identification analysis, bullet-lead comparison, and certain claims made by arson investigators. Other forensic areas of doubtful scientific value, but with experiential claims of reliability, include bitemark identification evidence and some additional areas of pattern-recognition expertise.²¹

¹⁹ For efforts to consider and taxonomize the functions of expertise and its use as evidence, and to identify distinct admissibility challenges posed by different kinds of evidence, see, e.g., D. Michael Risinger, *Preliminary Thoughts on a Functional Taxonomy of Expertise for the Post-Kumho World*, 31 SETON HALL L. REV. 508 (2000); Samuel R. Gross & Jennifer L. Mnookin, *Expert Information and Expert Evidence: A Preliminary Taxonomy*, 34 SETON HALL L. REV. 141 (2003); David L. Faigman, John Monahan & Christopher Slobogin, *Group to Individual (G2i) Inference in Scientific Expert Testimony*, 81 U. CHI. L. REV. 417 (2014)

²⁰ NAT’L RESEARCH COUNCIL OF THE NAT’L ACAD. OF SCIENCES, STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD (2009), <https://www.ncjrs.gov/pdffiles1/nij/grants/228091.pdf>.

²¹ PRESIDENT’S COUNCIL ON ADVISORS ON SCI. & TECH., REPORT TO THE PRESIDENT: FORENSIC SCIENCE IN CRIMINAL COURTS: ENSURING SCIENTIFIC VALIDITY OF FEATURE-COMPARISON METHODS 8 (2016), https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf.

Ultimately, experience only has value as a basis for grounding expert claims when there is a feedback loop that allows the expert to learn whether his or her experience is accurate. Hence, the harbor pilot's knowledge of obstructions is likely borne from actual feedback regarding their existence and danger, either from personal experience or the direct experience of his or her colleagues. In comparison, a forensic science such as hair identification analysis does not typically give the examiner any feedback on the accuracy of the exam. Such experts might learn that the defendant whose hair they examined was convicted or acquitted, but this provides at best limited information about the value or accuracy of their identification methods. Indeed, as we have seen with the long tenure that many claims based on experience have had, from blood-letting in medicine to bitemark identification in the law, self-described experts often remain convinced of the value of their expertise even in the face of contrary research data.

To be sure, in the case of medicine, art might indeed play something of a role in the treatment knowledge that doctors develop. After all, if a doctor has treated a certain illness numerous times, he or she is likely to learn what seems to work and what has not worked with similarly situated patients. For the most part—though not invariably—doctors receive feedback about the outcomes of their therapy. If the patient gets better, a doctor is likely to try the same treatment with the next similarly situated patient. Over time, continued success—or failure—with a treatment regimen will give the physician considerable useful, albeit informal, data on its utility. Yet, even on the treatment side of medicine, many doctors hold fast to treatments for a very long time that, ultimately, are demonstrated to be ineffective through careful research.²² Nonetheless, in regard to treatment outcomes, doctors are more like harbor pilots than hair identification analysts.

However, in *Wendell*, the question presented was not a treatment issue, nor a matter of diagnosis; it was a question of causation. On this question, doctors are unlikely to receive such direct feedback on their conjectures about what caused a particular illness. Feedback loops are frequently non-existent, or at least weaker, when physicians are assessing questions of medical causation.²³ Indeed, case law is replete with examples of speculative

²² See, e.g., David Epstein, *When Evidence Says No, But Doctors Say Yes*, ATLANTIC (Feb. 22, 2017), <https://www.theatlantic.com/health/archive/2017/02/when-evidence-says-no-but-doctors-say-yes/517368/> (arguing that although stents for stable patients provide no benefits, “hundreds of thousands of stable patients receive stents annually . . .”). See also Vinay Prasad et al., *A Decade of Reversal: An Analysis of 146 Contradicted Medical Practices*, 88 MAYO CLINIC PROC. 790 (Aug. 2013), <http://dx.doi.org/10.1016/j.mayocp.2013.05.012> (finding that 146 of 363 current medical practices are ineffective).

²³ To be sure, in some instances, physicians may have solid ways to test their theories of causation. If the physician's theory of cause is something testable—an allergen, or an exposure to something in the environment that can be controlled or tested, the physician may indeed receive feedback about what happens to the patient when that allergen is avoided. But

2018]

THE CURIOUS CASE

615

medical beliefs about causation that resulted in litigation only to turn out not to be valid when adequately researched. Bendectin, the subject of *Daubert* itself, and silicone implants, which generated hundreds of thousands of lawsuits, are two particularly prominent examples.²⁴ On issues of medical causation, then, medical doctors are more like hair identification analysts. They certainly might speculate about the causes of the illnesses that they treat, but the best—and only scientifically valid—answers to those questions will come from the research literature. And, ordinarily, this will involve some combination of toxicological and epidemiological studies.

B. *Relying on Qualifications Alone*

The fact that the Ninth Circuit was impressed with the qualifications of the two experts has marginal value to the review of the lower court's admissibility decision. The lower court also found plaintiffs' experts to be well-qualified clinicians. The issue was not whether they were impressive doctors; rather, it was the adequacy of the basis for their opinions regarding causation that the district court found wanting. In this respect, the Ninth Circuit fundamentally misunderstood the role of qualifications in the assessment of expert testimony. Rule 702, of course, requires that experts be qualified in their respective areas of expertise before being allowed to testify. But this is a necessary, not a sufficient, requirement. In fact, in the rule itself, qualifications operate as a prologue to the rule's substantial additional requirements: "a witness who is qualified as an expert by knowledge, skill experience, training or education may testify. . . if" (and, implicitly, *only* if) the evidence is: (a) helpful to the trier of fact; (b) based on sufficient facts and data; (c) the result of reliable principles and methods; and (d) reliably applied in this instance.²⁵ A surprising number of courts, however, appear to share the Ninth Circuit's overvaluing of this preliminary requirement.²⁶

for many diseases, the doctor will have no way to test a theory about cause at the level of the individual patient.

²⁴ JOSEPH SANDERS, *BENDECTIN ON TRIAL; A STUDY OF MASS TORT LITIGATION* (1998); MICHAEL D. GREEN, *BENDECTIN AND BIRTH DEFECTS: THE CHALLENGES OF MASS TOXIC SUBSTANCES LITIGATION* (1996); Laurens Walker & John Monahan, *Scientific Authority: The Breast Implant Litigation and Beyond*, 86 VA. L. REV. 801 (2000); MARCIA ANGELL, *SCIENCE ON TRIAL: THE CLASH OF MEDICAL EVIDENCE AND THE LAW IN THE BREAST IMPLANT Case* (1997).

²⁵ FED. R. EVID. 702.

²⁶ See, e.g., *Tampa Bay Water v. HDR Eng'g, Inc.*, 731 F.3d 1171, 1185 (11th Cir. 2013) (quoting *Quiet Tech. DC-8, Inc. v. Hurel-Dubois UK Ltd.*, 326 F.3d 1333, 1341 (11th Cir. 2003)) ("Although an expert's qualifications go primarily to the first prong of *Daubert's* inquiry, 'an expert's overwhelming qualifications may bear on the reliability of his proffered testimony' even if 'they are by no means a guarantor of reliability.'"). See generally *Graves v. Mazda Motor Corp.*, 405 F. App'x 296 (10th Cir. 2010) ("[T]he *ipse dixit* of an expert, no matter how qualified he may be, is never enough to guarantee him a ticket to admissibility.");

Beyond the plain meaning of Rule 702, there are many reasons why courts should not rest admissibility decisions on qualifications alone. Foremost, perhaps, is the significant danger, especially acute in our adversarial system, that even the most qualified professionals may be tempted to propound opinions that go beyond what sound science can support. The adversarial process already leads to the selection of party experts that are likely to be at the tails of scientific opinion²⁷ and case demands likely further prod experts to more categorical or extreme statements than they might otherwise maintain among their professional colleagues. Indeed, with all due respect to the plaintiffs' experts in *Wendell*, this is exactly what may have happened in that case. At least one of the experts explicitly conceded that his opinion "to a reasonable degree of medical certainty" for courtroom purposes would not meet the standards used for peer review and publication in science. Courts need to ensure that otherwise qualified experts stay within their expertise when testifying, particularly given the pressures of partisanship that may invite, consciously or not, overstatements that assist the party paying for their testimony.²⁸

Additionally, the entire notion of "qualifications" is a profession-specific classification.²⁹ A world-renowned oncologist might be extremely "well-qualified" as a treating physician but have scant expertise in identifying the causes of cancer. A DNA technician might be well qualified to perform the rote protocol of PCR testing, but lack expertise in understanding molecular biology and thus what the empirical basis for the test itself is. And a fingerprint expert might know how to apply the standard approach to identification—known as ACE-V³⁰—but have no understanding of its validity or lack thereof. A witness "qualified" in his or her field might,

DAVID L. FAIGMAN, EDWARD K. CHENG, JENNIFER L. MNOOKIN, ERIN E. MURPHY, JOSEPH SANDERS & CHRISTOPHER SLOBOGIN, *MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY* § 1:12 (2017–2018 ed. 2017).

²⁷ See Jonah B. Gelbach, *Expert Mining and Required Disclosure*, 81 U. CHI. L. REV. 131, 131 (2014) (describing "expert mining" as the practice, by "resourceful attorneys" of "hiring multiple experts, asking each to provide an expert report on the same issue, and then put[ting] on the stand only the one who provides the most favorable report").

²⁸ For an effort to develop institutional tools to combat this set of structural dilemmas, see JuriLytics, founded by David Faigman, one of this Article's authors. For the history of expert evidence showing how partisanship has been an extremely longstanding (and hard to combat) concern, see Jennifer L. Mnookin, *Idealizing Science and Demonizing Experts: An Intellectual History of Expert Evidence*, 52 VILL. L. REV. 763 (2007).

²⁹ See *RG Steel Sparrows Point, LLC v. Kinder Morgan Bulk Terminals, Inc.*, 609 F. App'x 731, 738 (4th Cir. 2015) ("The question of whether a witness is qualified to testify is context-driven and can only be determined by the nature of the opinion he offers.").

³⁰ ACE-V stands for Analysis, Comparison, Evaluation, Verification.

2018]

THE CURIOUS CASE

617

or might not, be “qualified” to testify on the specific issue in dispute at trial. In short, there is always a question of fit between an expert’s professional qualifications and whether he or she is qualified to provide the testimony offered in court.

C. *What is the “Task at Hand” in Medical Causation Testimony?*

A fundamental flaw in the Ninth Circuit’s review was its equating expertise in treating a disease with expertise in identifying the cause of that disease.³¹ These are substantially different skillsets and there is little basis for believing that clinical skill equates to scientific acumen. To be sure, some clinicians may also be accomplished research scientists, or at least competent interpreters of a research literature, but the fact that the Venn diagrams may well overlap does not mean that a qualified clinician should necessarily be permitted to testify about causation. Indeed, when the Ninth Circuit turned to the question of causation, it offered up both non sequiturs and inadequately supported conclusions.

Ordinarily, as the district court recognized, medical causation contains two separate levels of analysis, what is often referred to as general causation and specific causation. The former involves whether scientific support exists for the proposition that a particular drug or substance causes a particular illness; and the latter involves whether there is support for the proposition that a particular drug or substance caused the particular instance of that illness at issue in the case. General causation is a prerequisite to specific causation, since, if there is no proof that the drug or substance can cause the illness, then there can be no proof that it did cause the illness in a particular case.

Unfortunately, in *Wendell*, the Ninth Circuit did little more than restate the conclusory assertions of the plaintiffs’ experts, making no serious effort to evaluate their accuracy or basis. For example, Dr. Shustov, one of the plaintiffs’ two experts, said he relied “on medical records as well as [his] education, training and experience, knowledge of the pertinent medical literature and [his] knowledge of the epidemiology, diagnosis and natural history of HSTCL.”³² He said that he “pulled the facts out of the literature,” which indicated “an increased risk of HSTCL in patients taking 6-MP over the general population.”³³ Presumably, all of this was offered as a basis for a finding of general causation.

However, the Ninth Circuit did not provide any significant analysis of this literature, which, as the experts themselves conceded, lacked any

³¹ See *Wendell v. GlaxoSmithKline LLC*, 858 F.3d 1227, 1233–34 (2017).

³² *Id.* at 1234.

³³ *Id.*

toxicological or epidemiological studies. From what the Ninth Circuit discusses in the four corners of its opinion, it is challenging to understand what to make of the claim of an “increased risk” in this context, since they described no effort to control for confounding variables. Such statements are reminiscent of the spurious causal connection between the amount of ice cream consumed at a beach and the number of drownings; they are, of course, associated, but ice cream consumption in no way causes drownings: they are linked by a third variable, the number of beachgoers, and perhaps also by a fourth, the warmth of the day, which likely affects both how many beachgoers take to the water and how many go for the ice cream. This example, and Dr. Shustov’s opinion, both risk the elementary error of assuming that correlation demonstrates causation. The Ninth Circuit offered no analysis to establish that Dr. Shustov’s opinion was anything more than superficially plausible conjecture and speculation, grounded in anecdote, and dressed up in the garb of scientific jargon. Dr. Shustov claimed that, “[a]fter reviewing the literature, he ‘compiled the numbers about frequency of diseases, about frequency of inflammatory bowel disease and [he] looked at the biological causation of lymphoma pertaining to this case.’”³⁴ But what does this last sentence even mean without further explication? The Ninth Circuit made no attempt to explain it.

After apparently reaching a conclusion about the general causation between 6-MP and HSTCL, Dr. Shustov went on to offer an opinion on specific causation. In medical causation cases, this is typically achieved through a method best described as “differential etiology.” This form of analysis requires first ruling in the putative cause (here 6-MP) as possible and then ruling out other possible causes as inapplicable to the facts. This notion of differential etiology is sometimes confused with a similar worded method called “differential diagnosis.” Differential diagnosis is a method for identifying what illness a person suffers from; differential etiology, in contrast, is a method for determining what caused that illness. The two sound similar, but require very different knowledge bases and skillsets.

According to Dr. Shustov, however, the two are essentially the same. He said that “he performs differential diagnosis in attempting to diagnose every patient, and that he has applied the same technique to determine the cause of a disease.”³⁵ According to the court, his differential [etiology] “assumes the pertinence of all potential causes, then rules out the ones as to which there is no plausible evidence of causation, and then determines the most likely cause among those that cannot be excluded.”³⁶ The Ninth Circuit

³⁴ *Id.*

³⁵ *Id.*

³⁶ *Id.*

2018]

THE CURIOUS CASE

619

found this to be “scientifically sound.”³⁷ But given that there was insufficient proof of general causation, and that over 70% of HSTCL cases are idiopathic—that is to say, without known cause—the soundness of this differential analysis is far from obvious. The court summarized Dr. Shustov’s reasoning regarding his conclusion on the cause of Maxx’s illness as follows:

Dr. Shustov stated that there was a one in six million chance that Maxx would have developed HSTCL without being exposed to 6-MP. In light of those odds, Dr. Shustov stated that “based on [his] experience in T-cell lymphomas, knowledge of the literature and being involved in T-cell lymphoma research in the past ten years” he determined “that it’s much more likely that exposure to mutagen and immunosuppressants caused the lymphoma.”³⁸

The scientific logic of this analysis is, at a minimum, inadequately specified, and may well be far from sound. Ultimately, the court returned to the qualifications of the two experts, once again conflating their medical credentials with the soundness of their testimony. The court asserted, “[n]othing in *Daubert*, or its progeny, properly understood, suggests that the most experienced and credentialed doctors in a given field should be barred from testifying based on a differential diagnosis.”³⁹ But, in fact, everything in *Daubert* and its progeny suggests that “the most experienced and credentialed doctors in a given field should be barred from testifying” if their differential diagnosis is not sufficiently scientifically valid.⁴⁰

In addition to its emphasis on credentials as justifying the legitimacy of the experts’ conclusions, the appellate court seemed to think, more generally, that the district court’s approach to evaluating the experts was too formalistic and cramped. The Ninth Circuit wrote,

The district court looked too narrowly at each individual consideration, without taking into account the broader picture of the experts’ overall methodology. It improperly ignored the experts’ experience, reliance on a variety of literature and studies, and review of Maxx’s medical records and history, as well as the fundamental importance of differential diagnosis by experienced doctors treating troubled patients. The district court also overemphasized the facts that (1) the experts did not develop their opinions based on independent research and (2) the experts did not

³⁷ See *Wendell*, 858 F.3d at 1234.

³⁸ *Id.*

³⁹ *Id.* at 1235.

⁴⁰ *Id.*

cite epidemiological studies. We hold that all together, these mistakes warrant reversal.⁴¹

The appellate court is asserting, in essence, that the district court's analysis was excessively atomistic, "look[ing] too narrowly at each individual consideration" rather than "the broader picture. . . ."⁴² On this point, the Ninth Circuit's admonition is correct in theory but deeply problematic in application.

It is true that some courts have been excessively atomistic in their analysis of expert claims. Excessive atomism risks making two errors, one evidentiary and the other scientific.⁴³ As a matter of evidence law, "a brick is not a wall"—that is, proof of a necessary element can be made up of many distinct bricks, and no individual item of evidence must prove the point alone. From an evidentiary perspective, there is nothing inappropriate about establishing causation by aggregating multiple items of evidence, no one of which establishes causation alone. And scientists, too, can and do aggregate multiple discrete and disparate items of evidence to reach a conclusion. Some scientific aggregation methods may be formal and methodologically rigorous, like structured meta-analyses, while others may be more informal or the product of collective engagements by experts, like the Cochrane reviews, or government consensus panels, or assessing general causation by using the Bradford Hill guidelines.⁴⁴ When aggregation is legitimate, and when it may become an excuse for insufficiently justified expert conclusions based on "soi-disant" expertise and hand-waving is not simple to answer. It depends, both, on the details of the underlying evidentiary support for a given claim, as well as the specific question being asked. This focus on whether the scientific basis supports the particular, concrete testimony offered in court was referred to as the "task at hand" in *Kumho Tire v. Carmichael*, the third case in the *Daubert* trilogy, and is an important concept defined and

⁴¹ *Id.* at 1233.

⁴² *Id.*

⁴³ Jennifer L. Mnookin, *Atomism, Holism, and the Judicial Assessment of Evidence*, 60 UCLA L. REV. 1524, 1576–80 (2013).

⁴⁴ See generally Lawrence Cetrulo, *Evaluating the Totality of the Evidence: The Bradford Hill Criteria*, in 3 TOXIC TORTS LITIG. GUIDE § 22:5 (2017); Kirsten Bell, *Cochrane Reviews and the Behavioural Turn in Evidence-Based Medicine*, 21 HEALTH SOC. REV. 313 (2012); David C. Currow & Irene J. Higginson, *Cochrane Reviews: Four Proposals for Improvement*, 18 J. PALLIATIVE MED. (2015) 906–07; Matthew J. Gabel & Charles R. Shipan, *A Social Choice Approach to Expert Consensus Panels*, 23 J. HEALTH ECON. 543–64 (2004); Dr. Frank C. Woodside and Allison G. Davis, *The Bradford Hill Criteria: The Forgotten Predicate*, 35 T. JEFFERSON L. REV. 103 (2013); P.M. Wortman, *Consensus Panels: Methodology*, INT'L ENCY. SOC. BEHAV. SCI. 2609–13 (2001); COCHRANE HANDBOOK FOR SYSTEMATIC REVIEWS OF INTERVENTIONS (Julian P.T. Higgins & Sally Green eds. 2011), <http://handbook-5-1.cochrane.org/>.

refined brilliantly by Professor Michael Risinger in his scholarship.⁴⁵

At one extreme, there is a danger of courts applying *Daubert* too rigidly, piece by piece, and looking too narrowly at each item of evidence—perhaps even insisting upon a single “smoking gun” piece of causation evidence, like a precisely on point epidemiological study showing a substantially heightened relative risk, as a necessary precondition to proving cause. An excessively atomistic judicial analysis pulls apart each item of evidence, dismissing each strand, study, or claim as inadequately establishing what needs to be proven, providing inadequate opportunity to assess whether the evidence, taken in its entirety, makes the case that the “weight of the evidence” establishes causation by a preponderance.⁴⁶

But in this case, the Ninth Circuit teeters distressingly far in the other direction. The opinion’s analysis emphasizes—about a number of issues and concerns—that each is not necessary for evidence to be admissible under *Daubert* and Rule 702.⁴⁷ On each of these points taken individually, the Ninth Circuit is correct. But in aggregate, it errs, particularly given the alleged application of *Joiner’s* abuse of discretion standard, by failing to recognize that while *any* given limitation might not be fatal, to have so *many* strongly suggests that there are not adequate scientific grounds that support the opinion, or at a minimum, that the district court’s determination was squarely within the bounds of reasoned discretion.

For example, as the court correctly notes, it ought not necessarily to be fatal that an expert’s opinion does not derive from independent research.⁴⁸ While opinions derived specifically for court escalate the danger of partisanship, often questions come to the fore as a result of litigation and so the fact that experts developed their opinions in that posture ought not, by itself, preclude admissibility.⁴⁹ Similarly, epidemiological studies, while enormously helpful, ought not to be a *sine qua non*, particularly because for

⁴⁵ *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 141 (1999); D. Michael Risinger, *Defining the “Task at Hand”: Non-Science Forensic Science After Kumho Tire Co. v. Carmichael*, 57 WASH. & LEE L. REV. 767 (2000).

⁴⁶ The legitimacy of “weight of the evidence” approaches to assessing toxic torts cases has been debated by both courts and commentators. See, e.g., *Milward v. Acuity Specialty Prods. Grp., Inc.*, 639 F.3d 11 (1st Cir. 2011); Mnookin, *supra* note 42; Michael D. Green, *Pessimism about Milward*, 3 WAKE FOREST J. L. & POL’Y 41 (2013); Sheldon Krinsky, *The Weight of Scientific Evidence in Policy and Law*, 95 AM J. PUB. HEALTH S129 (2005), <http://ajph.aphapublications.org/doi/pdf/10.2105/AJPH.2004.044727>.

⁴⁷ *Wendell v. GlaxoSmithKline LLC*, 858 F.3d 1227, 1234-36 (9th Cir. 2017).

⁴⁸ *Id.* at 1235.

⁴⁹ See *Daubert v. Merrell Dow Pharm., Inc.*, 43 F.3d 1311, 1316–19 (9th Cir. 1995) (emphasizing on remand the significance of this criterion in assessing expert testimony); Jennifer L. Mnookin, *Expert Evidence, Partisanship, and Epistemic Competence*, 73 BROOK. L. REV. 1009, 1014, 1023 (2008). See also David Sonenshein & Charles Fitzpatrick, *The Problem of Partisan Experts and the Potential for Reform through Concurrent Evidence*, 32 REV. LITIG. 1 (2013).

truly rare, low incident diseases, they may be wholly implausible, and, more generally, because it may be possible to aggregate non-epidemiological evidence in a way that supports a finding of both general and specific causation by a preponderance of the evidence. A lack of animal studies and toxicological assessments also ought not necessarily be fatal either, if other strong evidence (like, perhaps, well-designed epidemiological studies) supports causation. The same goes for a lack of a well-specified biological mechanism for causation—again, if other persuasive evidence, like strong epidemiological support, exists, then an absence of a detailed or proven causal mechanism may be acceptable. But when *all* of these are absent—no animal studies, no epidemiological evidence, no causal mechanism beyond the fact that one of the drugs is known to be carcinogenic—and when the cancer itself is largely idiopathic, then credentials, conjectures and case reports cannot simply be alchemically combined to produce valid scientific conclusions.

It may be that a particular set of facts and circumstances could, in rare instances, warrant a finding of causation even without any of these traditional hallmarks of scientific causation—but in that case, one would want to see very carefully reasoned, thoughtful engagement of how and why causation could nonetheless be inferred in this “task at hand,” from whatever evidence in fact supported it. In *Wendell*, however, the Ninth Circuit did not offer such an analysis.

Given the absence of toxicology, epidemiology, or a clear biological mechanism, why did the Ninth Circuit strain so hard to overturn the summary judgment ruling of the district court, notwithstanding the deferential abuse of discretion standard? Why did they engage in such a dramatically non-stringent approach to *Daubert* and Rule 702, overturning the trial court’s ruling and deeming admissible expert testimony with distressingly little significant scientific support?

Although they did not offer a detailed assessment of what persuaded them, part of the answer likely lies in the superficial plausibility of the case reports in this particular area, even standing alone. HSTCL is extremely rare, with only a total of a few hundred reported cases worldwide since the disease variant was identified two decades ago. And a number of these cases have in fact been reported in young men with conditions quite like Maxx’s—indeed a 2013 aggregation of case reports found thirty-seven cases of HSTCL in patients with medical conditions similar to Maxx’s, and in three-quarters of those cases, the patients had taken a pair of medicines similar to Maxx’s.⁵⁰ Most of these cases occurred in younger men – again, like Maxx.

⁵⁰ Saranya A. Selvaraj et al., *Use of Case Reports and the Adverse Event Reporting System in Systematic Reviews: Overcoming Barriers to Assess the Link Between Crohn’s Disease Medications and Hepatosplenic T-Cell Lymphoma*, 2:53 SYSTEMATIC REV. 1 (2013),

Indeed, because of the reporting of HSTCL in the adverse event reporting system for approved drugs, these medicines now contain warnings that alert physicians and patients to the possibility of an elevated risk of HSTCL. Given the rarity of the disease, the number of HSTCL cases diagnosed in patients on these kinds of drugs certainly looks like it might be more than a coincidence. But as the 2013 publication asserts, the authors cannot establish “a causative effect other than ‘possible’” because of “the limited applicability of causality assessment tools for rare irreversible events.”⁵¹ “Possible,” standing alone, is not enough to establish legal causation by a preponderance of the evidence.

Although the 2013 aggregation cited above was not referenced by the experts, one of the experts in the case had, in fact, relied on an earlier published aggregation of case reports that also illustrated a substantial “cluster” of thirty-six cases of HSTCL in patients similar to Maxx.⁵² Of the approximately 200 cases of HSTCL reported worldwide since this variant of lymphoma was identified in the mid-90s, thirty-six of them are associated with IBD patients receiving thiopurines, and twenty of those were also receiving some anti-TNF therapy as well. The vast majority of those IBD patients diagnosed with HSTCL were, like Maxx, young and male.

This is striking data. For an extraordinarily rare disease of unknown cause, the fact that approximately twenty percent of the known instances are associated with both a relatively common illness and a particular family of drug treatments is certainly suggestive. The authors of the 2011 article offer some slightly back-of-the-envelope relative risks based on the data they have available, which suggest a substantially higher risk of HSTCL among men younger than thirty-five years old exposed either to thiopurines, or to both thiopurines and anti-TNF medications, than among IBD patients generally.⁵³

To infer causation from case reports is, to say the least, fraught, but the apparent extreme rarity of the disease overall combined with a significant minority of those cases being linked to treatments similar to Maxx’s was what led the plaintiffs’ experts to assert their belief in causation “to a reasonable degree of medical probability.” As Dr. Andrei Shustov wrote in his report in the case:

<https://systematicreviewsjournal.biomedcentral.com/track/pdf/10.1186/2046-4053-2-53?site=systematicreviewsjournal.biomedcentral.com>.

⁵¹ *Id.*

⁵² Expert Report & Affidavit of Andrei Shustov, *Wendell v. Johnson & Johnson*, No. 09-cv-04124, 2014 WL 10519380 (N.D. Cal. Jan. 31, 2014) (citing David S. Kotlyar et al., *A Systematic Review of Factors That Contribute to Hepatosplenic T-Cell Lymphoma in Patients with Inflammatory Bowel Disease*, 9 *CLINICAL GASTROENTEROLOGY & HEPATOLOGY* 36 (2011), [http://www.cghjournal.org/article/S1542-3565\(10\)00907-9/pdf](http://www.cghjournal.org/article/S1542-3565(10)00907-9/pdf)). (Note that many of these cases are the same cases of HSTCL as those in the other study.)

⁵³ *Id.*

Given the absolute rarity of this disease generally, a cluster of 36 cases arising in young, predominantly male patients treated for IBD with thiopurines and TNF antagonists stands as almost a signature of the disease. While the precise mechanism by which these drugs used in the setting of IBD in young patients give rise to HSTCL is not known, it is clear that the use of these drugs, either individually (in the case of the thiopurines) or in combination, either causes or contributes to the development of HSTCL in certain patients. This high incidence of an exceedingly rare cancer in this distinct cohort is compelling evidence of causation.

Maxx Wendell was one of those patients. He was a young male with ulcerative colitis (a form of inflammatory bowel disease) with a history of 5+ years of treatment with a thiopurine in combination with the TNF antagonists Remicade and Humira who developed an exceedingly rare cancer almost uniquely associated with this treatment regimen in this cohort. To a reasonable degree of medical probability, the combination use of a thiopurine with TNF antagonists for the treatment of his inflammatory bowel disease caused, or substantially contributed, to the development of HSTCL to which he succumbed four months after diagnosis despite multiple aggressive therapies.

I hold all of these opinions to a reasonable degree of medical probability.⁵⁴

Interestingly, also in 2013, Dr. Shustov wrote a blog post about HSTCL, for the TCLLF, the T-Cell Leukemia Lymphoma Foundation. There he couched his views in far less certain terms: “No specific cause of HSTCL has been identified so far. However, in some cases, long immunosuppression has been implicated. There is a suggestion that young people who were treated for childhood inflammatory bowel disease (such as ulcerative colitis and Crohn’s disease) might be at risk for developing HSTCL.”⁵⁵

The space between referencing “a suggestion” that treatments like Maxx’s “might” be implicated in causing HSTCL and asserting “compelling evidence of causation” to a “reasonable degree of medical probability” is

⁵⁴ Expert Report & Affidavit of Andrei Shustov, *Wendell v. Johnson & Johnson*, No. 09-cv-04124, 2014 WL 10519380 (N.D. Cal. Jan. 31, 2014).

⁵⁵ Andrei Shustov, *Hepatosplenic T-Cell Lymphoma (HSTCL)*, T-CELL LEUKEMIA LYMPHOMA FOUND. (Dec. 25, 2015), <http://www.tcllfoundation.org/blog/learning-corner/learning-corner/hepatosplenic-t-cell-lymphoma-hstcl>.

fairly substantial. Obviously, these were written for different audiences, and perhaps it is unfair to compare the language used in these two settings to one another. But it is, in any event, worth detailing why the “cluster” evidence, while dramatic, may not be as persuasive as it seems to Dr. Shustov (and perhaps to the Ninth Circuit). First, IBD itself is extremely common—in the United States and Europe, the authors of the 2011 review estimate that 3.6 million people have IBD. Many of those—in one study, roughly forty-four percent—have been exposed to thiopurines, and only a tiny handful have gotten this rare form of cancer. That of course does not diminish the suggestion that risk may be elevated with exposure, but it does make the existence of confounds or other causes a matter of serious question. Moreover, the fact that among IBD-linked cases, HSTCL seems to develop predominantly in young men raises questions. Might there be something about these men’s IBD itself, or some other genetic propensity, that caused their HSTCL? It is important to recognize that these case reports could not exclude that possibility. Furthermore, information from the adverse reporting systems for drugs is understood to be far from perfect, so the underlying data upon which these analysts were relying may have significant weaknesses. In addition, notwithstanding Shustov’s assertion to the contrary—and as pointed out explicitly by the defense experts—the association of HSTCL with young men with IBD and a certain treatment regimen did not amount to a “signature.” (Occasionally some diseases are so exclusively associated with a given exposure that they amount to a signature, like the association of asbestosis with exposure to asbestos.) But recall that most known HSTCL cases are in fact idiopathic; that fact alone establishes that HSTCL is therefore not a signature disease for IBD plus a specific drug regimen. Moreover, some HSTCL cases have occurred in non-IBD patients after organ transplants (some of whom also had taken other immunosuppressants, though different ones); as well as in some patients with different diseases, some with no other diseases, and sometimes in women and in older patients.

Finally, and critically, and as the defense experts emphasized in their affidavits, case reviews simply are not valid science.⁵⁶ As one of the defense experts wrote, “While case reports may be suggestive of possible problems that are worthy of further study, they in and of themselves do not constitute proof of cause and effect relationships due to their anecdotal, potentially confounded, non-standardized, and totally uncontrolled nature.”⁵⁷ The other defense expert concurs, claiming that “no reliable scientific evidence”

⁵⁶ See, e.g., Expert Report & Affidavit of Robert J. Valuck, *Wendell v. Johnson & Johnson*, No. 09-cv-04124, 2013 WL 11025314 (N.D. Cal. Sept. 19, 2013); Expert Report & Affidavit of Andrew Place, No. 09-cv-04124, 2014 WL 10519379 (N.D. Cal. Jan. 9, 2014).

⁵⁷ *Id.*

supports causation.⁵⁸

Without doubt, much about HSTCL and its causes is unknown. What we have then, is a battle of the experts in a condition of genuine scientific uncertainty. For the plaintiffs' experts, the cluster of known cases and the association of similar drug regimens with HSTCL in patients with demographic similarities to Maxx is so striking, given the extreme disease rarity, that they are willing to name his treatment regimen as the cause of his cancer, even in the absence of the kinds of evidence one would typically wish for to make such a judgment. By contrast, the defense experts emphasize that the risk assessments within these aggregated case reports amount to guesswork and conjecture, not science—at best, a hypothesis rather than proof. From their perspective, we currently have, in fact, a complete lack of epidemiological investigation, assessments using thoughtful case controls, or any other genuine scientific basis for reaching a conclusion about causation or making an accurate “relative risk” judgment about this medication and this disease.

If we tamp down the slightly overwrought language on both sides, it may be that both perspectives are at least partly right. What we have is a plausible hypothesis of a causal link that has some meaningful, albeit imperfect and limited, evidentiary support. But we also have, as the defense recognized, a significant lack of the forms of scientific evidence that we would typically deem necessary to assert causation.

There are, to be sure, potential public policy justifications for permitting cases like Maxx's to go to a factfinder even without what we would generally wish for in terms of an adequate scientific foundation. For rare, infrequent diseases like HSTCL, we may never have enough solid scientific evidence to establish cause—and this may generate a structural risk of under-deterrence if we insist upon solid science as a prerequisite to getting to trial. Sometimes, the drugs at issue really will have caused the harm, while under our current rules, plaintiffs may never be able to adequately prove this causation with epidemiology or a strong understanding of the underlying mechanism, especially when the disease itself is extremely rare. (Of course, the opposite danger of over-deterrence may be at least as significant. If we revise our tort law,⁵⁹ or shift burdens of proof,⁶⁰ or lower our expert evidence standards so that manufacturers face trial notwithstanding shaky and limited evidence supporting causation, valuable products may become costlier or

⁵⁸ Expert Report & Affidavit of Andrew Place, No. 09-cv-04124, 2014 WL 10519379 (N.D. Cal. Jan. 9, 2014).

⁵⁹ See, e.g., Margaret A. Berger, *Eliminating General Causation: Notes Towards a New Theory of Justice and Toxic Torts*, 97 COLUM. L. REV. 2117 (1997).

⁶⁰ See, e.g., *Zuchowicz v. United States*, 140 F.3d 381 (2d Cir. 1998); Kenneth S. Abraham, *Self-Proving Causation*, 99 VA. L. REV. 1811 (2013).

2018]

THE CURIOUS CASE

627

altogether kept from the market and from consumers who could benefit from them.)

Legal decision-making under conditions of genuine scientific uncertainty thus raises challenging issues, and *Wendell* is a prime example. But we would suggest that we should face these public policy issues—and these questions of tort law—squarely, rather than straining our rules of evidence and treating expert conjecture as if it were science.

IV. CONCLUSION

In this Article, we have sought to use the curious case of *Wendell* to illustrate some of the fundamental challenges associated with assessing causation in toxic tort cases. Why, precisely, have we called *Wendell* a curious case? First, because we find the Ninth Circuit opinion very hard to reconcile with their supposedly deferential standard of review. The district court was almost certainly well within its discretion to exclude the proffered expert testimony and the circuit court applied, it would seem, some level of heightened scrutiny to reverse that decision. Moreover, the Ninth Circuit offered no substantive explanation for its disagreement with the lower court, outside of its excessive—and curious—deference to the experts' clinical credentials. We think this credential-centric approach to the assessment of expert evidence is not merely curious, but mistaken as both a matter of science and as a matter of law.

It is critical to recognize, however, that *Wendell* is *not* curious in the sense that the challenges it poses are unique or unusual: assessing whether legally sufficient evidence supports causation is required in every toxic torts case. Today, nearly 25 years after *Daubert*, courts still struggle with the task; hence, we believe, this exploration of *Wendell* offers valuable lessons that go well beyond the case. *Wendell* is also not curious in representing—albeit in extreme form—the difficulty of drawing categorical distinctions between admissibility and exclusion in causation contexts in which little research is available. Especially when the scientific record is thin, the fine line between legitimate inference from incomplete evidence, and inappropriate speculation and conjecture, becomes both absolutely critical and particularly challenging to navigate. Because many of the dynamics present in *Wendell* are not uncommon, it is worth concluding with a few brief points about the case and how, in our view, it should have been resolved.

- (1) First, we strongly believe that if “abuse of discretion” is indeed the appropriate standard of review, as *G.E. v. Joiner* clearly stated, then the trial court's determination ought not to have been overturned. We recognize that the cluster of case reports is suggestive of the possibility of a causal link, but it was

entirely reasonable for the trial judge to decide that suggestive though they were, they were not sufficient to support the plaintiffs' experts' opinions about causation, given the near-total absence of more traditional scientific evidence and studies. (There are, in our view, some solid arguments in favor of *de novo* review for expert claims that go beyond the individual case, like general causation, but this debate goes beyond the scope of our discussion here. But under *G.E. v. Joiner*, we believe that the Ninth Circuit itself erred when it asserted that the district court's exclusion of the plaintiffs' experts constituted legal error, given both the record and the district court's analysis of it.)

- (2) Second, even if we imagine that the Ninth Circuit had been determining admissibility under a *de novo* standard, we are disappointed by the reasoning of the court and its strong reliance on the experts' credentials. It is critical for judges assessing expert evidence under *Daubert* to go beyond the expert's bona fides and assess the expert's claims and whether they have adequate scientific or epistemic support. Credentials need to be a starting point for an analysis of the admissibility of expert evidence, not the heart of it, and certainly not a justification for a qualified expert to offer opinions insufficiently based on reliable methods and knowledge.
- (3) That said, if we imagine that they had been analyzing the case under a *de novo* standard, we believe that the Ninth Circuit could have legitimately ruled in favor of the admissibility of the plaintiffs' evidence. To do so should have necessitated a careful, critical look at the disease cluster evidence, and a thoughtful engagement of why, given the extraordinary rarity of the disease and the substantial fraction of cases apparently associated with drug regimens much like Maxx's, an inference of causation, while far from certain, arguably could meet the preponderance standard. Such an argument might also usefully engage in careful thinking about what the legal system should do when the evidence we would like to have may well never exist. Should we think about expert evidence differently when the epidemiological study we might wish for simply happens not to have been done, compared to those instances when the rarity of the disease at issue makes it that much harder to study? Should we build some kind of "necessity" or "best realistic

2018]

THE CURIOUS CASE

629

evidence” standard into the evidence rules, and if so, how ought it to be operationalized?⁶¹

- (4) Our analysis suggests that in assessing courts’ actions under *Daubert*, it is *both* the conclusion and the reasoning that matters. If the trial court had deemed the plaintiffs’ experts’ testimony admissible on the grounds that the experts were adequately credentialed, that would, in our view, be an abuse of discretion. If, however, the trial court deemed the plaintiffs’ experts’ testimony admissible on the basis of a thoughtful assessment that explained why the inference of causation, while not overwhelming, was adequate given the task at hand and the facts available, the same conclusion would not be an abuse of discretion. We do not think that appellate courts need any form of “mixed” standard of review to make and apply this distinction appropriately; the simple point is that in assessing whether the district court exercised its discretion in a reasonable way, the reasons given and the quality of the analysis matter.
- (5) Finally, this analysis suggests that at the district court level, properly assessed, either admissibility or exclusion could have been legitimate. Perhaps this is surprising, or even troubling, given that we are talking about a scientific question of causation. But it is, in fact, simply the consequence of (a) a heightened admissibility standard like *Daubert*, combined with (b) a flexible, multi-pronged standard-rather-than-rule approach to reliability, and (c) an abuse of discretion standard.

Perhaps *Wendell* can best be seen as a case that helps to establish that old adage that hard cases make bad law. The appellate judges may have had the inchoate sense that the substantial cluster of similar cases, coupled with the rarity of the disease, made the physicians’ claims of causation credible. But the opinion took two important wrong turns: first, while nominally applying an abuse of discretion standard, it in fact gave little discretion to the trial judge’s quite reasonable conclusions. Second, the opinion placed far too much emphasis on the credentials of the experts, rather than carefully assessing the substantive basis for the causation claim. To be sure, assessing

⁶¹ This is an issue we have each wrestled with elsewhere. See David L. Faigman et al., *How Good is Good Enough?: Expert Evidence Under Daubert and Kumho*, 50 CASE W. RES. L. REV. 645, 665 (2000); Jennifer L. Mnookin, *The Courts, the NAS, and the Future of Forensic Science*, 75 BROOK. L. REV. 1209 (2010).

the merits of the relevant scientific claims is a substantially harder task than assessing the experts' credentials. In dissent in *Daubert*, Justice Rehnquist worried about whether the courts were up to the task that *Daubert* set for them: "I do not doubt that Rule 702 confides to the judge some gatekeeping responsibility in deciding questions of the admissibility of proffered expert testimony. But I do not think it imposes on them either the obligation or the authority to become amateur scientists in order to perform that role." We do not believe that to fulfil their duties under *Daubert*, judges assessing causation must become full-fledged amateur scientists. But we do believe that they absolutely must delve into the substance of the scientific evidence. They need to avoid arguments based on shortcuts like a near-exclusive focus on credentials and must instead ground their conclusions upon careful, case-specific assessment of the adequacy of the scientific and empirical evidence. We grant, as Justice Rehnquist intimated, that this is no easy task, but we have every confidence that the judiciary has not only the obligation but the ability to do so.