

United States Court of Appeals For the First Circuit

No. 09-2270

BRIAN K. MILWARD AND LINDA J. MILWARD,

Plaintiffs, Appellants,

v.

ACUITY SPECIALTY PRODUCTS GROUP, INC., ET AL.,

Defendants, Appellees.

APPEAL FROM THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MASSACHUSETTS

[Hon. George A. O'Toole, Jr., U.S. District Judge]

Before

Lynch, Chief Judge,
Lipez and Howard, Circuit Judges.

Steve Baughman Jensen, with whom Allen Stewart, P.C., James Gotz, and Kreindler & Kreindler were on brief, for appellants.

Raphael Metzger, Gregory Coolidge, and Metzger Law Group on brief for the Council for Education and Research on Toxins, et al., amici curiae.

Joseph J. Leghorn, with whom Nixon Peabody LLP was on brief, for appellees.

March 22, 2011

LYNCH, Chief Judge. Brian and Linda Milward brought negligence claims against defendant chemical companies alleging that the rare type of leukemia that Brian Milward suffers, Acute Promyelocytic Leukemia (APL), was caused by his routine workplace exposure to benzene-containing products that had been manufactured or supplied by defendants. Milward worked as a refrigeration technician and asserted that he was exposed to benzene from 1973 until the time he filed this complaint and jury demand in October 2007. He had been diagnosed with APL in October 2004.

At defendants' request, the district court bifurcated the suit into two phases. The first phase concerned whether the expert opinion offered by plaintiffs on "general causation" was admissible under Federal Rule of Evidence 702. "'General causation' exists when a substance is capable of causing a disease." Restatement (Third) of Torts: Liability for Physical and Emotional Harm § 28 cmt. c(3) (2010) ("Restatement"). If plaintiffs' expert evidence had been ruled admissible, the second phase would have considered all other issues, including negligence, exposure, and the "specific causation" of Milward's leukemia. "'Specific causation' exists when exposure to an agent caused a particular plaintiff's disease." Id. § 28 cmt. c(4).

This case never reached the second phase. The district court ruled that the testimony of plaintiffs' expert on general causation, Dr. Martyn Smith, was inadmissible under Federal Rule of

Evidence 702. The court so ruled after reviewing written statements and materials and conducting a four-day evidentiary hearing in which it heard testimony from plaintiffs' experts Dr. Smith, a toxicologist, and Dr. Carl Cranor, an expert on scientific methodology; and from defendants' experts Dr. David Garabrant, an epidemiologist, Dr. David Pyatt, a toxicologist, and Dr. John Bennett, a pathologist. The district court, in a detailed opinion, ruled that "Dr. Smith's proffered testimony that exposure to benzene can cause APL lacks sufficient demonstrated scientific reliability to warrant its admission under Rule 702." Milward v. Acuity Specialty Prods. Grp., Inc., 664 F. Supp. 2d 137, 140 (D. Mass. 2009). The court entered final judgment for defendants and plaintiffs timely appealed.

The appellate standard of review for Rule 702 rulings is abuse of discretion. Gen. Elec. Co. v. Joiner, 522 U.S. 136, 146 (1997). "This standard is not monolithic: within it, embedded findings of fact are reviewed for clear error, questions of law are reviewed de novo, and judgment calls are subjected to classic abuse-of-discretion review." Ungar v. Palestine Liberation Org., 599 F.3d 79, 83 (1st Cir. 2010); see also Baker v. Dalkon Shield Claimants Trust, 156 F.3d 248, 251-52 (1st Cir. 1998) (noting these three dimensions of the abuse of discretion standard in reviewing exclusion of expert testimony).

We reverse the district court's exclusion of Dr. Smith's general causation testimony. Cf. Ruiz-Troche v. Pepsi Cola of P.R. Bottling Co., 161 F.3d 77 (1st Cir. 1998) (reversing exclusion of expert testimony); Dalkon Shield, 156 F.3d 248 (same). Dr. Smith's testimony is admissible. We stress that it is up to the jury to decide whether to accept his opinion that exposure to benzene can cause APL--a proposition that plaintiffs must prove by a preponderance of the evidence.

I.

The Supreme Court in Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993), vested in trial judges a gatekeeper function, requiring that they assess proffered expert scientific testimony for reliability before admitting it.¹ The Court held that Rule 702 displaced the "general acceptance" test of Frye v. United States, 293 F. 1013 (D.C. Cir. 1923), under which "the admissibility of an expert opinion or technique turned on its 'general acceptance' vel non within the scientific community." Ruiz-Troche, 161 F.3d at 80. Under Rule 702:

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, if (1) the testimony

¹ Kumho Tire Co. v. Carmichael, 526 U.S. 137 (1999), clarified that courts have this function with respect to all expert testimony, not just scientific.

is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case.

Fed. R. Evid. 702.

The Daubert Court identified four factors that might assist a trial court in determining the admissibility of an expert's testimony: "(1) whether the theory or technique can be and has been tested; (2) whether the technique has been subject to peer review and publication; (3) the technique's known or potential rate of error; and (4) the level of the theory or technique's acceptance within the relevant discipline." United States v. Mooney, 315 F.3d 54, 62 (1st Cir. 2002) (citing Daubert, 509 U.S. at 593-94). These factors "do not constitute a 'definitive checklist or test.'" Kumho Tire Co. v. Carmichael, 526 U.S. 137, 150 (1999) (emphasis omitted) (quoting Daubert, 509 U.S. at 593). Given that "there are many different kinds of experts, and many different kinds of expertise," these factors "may or may not be pertinent in assessing reliability, depending on the nature of the issue, the expert's particular expertise, and the subject of his testimony." Id.

Exactly what is involved in "reliability" was not and could not have been filled out by Daubert. Rather, the answers must come from developing case law in adjudicating individual controversies. "[T]he question of admissibility 'must be tied to the facts of a particular case.'" Beaudette v. Louisville Ladder,

Inc., 462 F.3d 22, 25-26 (1st Cir. 2006) (quoting Kumho Tire, 526 U.S. at 150).

Although Daubert stated that trial courts should focus "on principles and methodology, not on the conclusions that they generate," Daubert, 509 U.S. at 595, the Court subsequently clarified that this focus "need not completely pretermit judicial consideration of an expert's conclusions," Ruiz-Troche, 161 F.3d at 81 (citing Joiner, 522 U.S. at 146). In Joiner, the Court explained that "conclusions and methodology are not entirely distinct from one another" and "nothing in either Daubert or the Federal Rules of Evidence requires a district court to admit opinion evidence that is connected to existing data only by the ipse dixit of the expert." Joiner, 522 U.S. at 146. Expert testimony may be excluded if there is "too great an analytical gap between the data and the opinion proffered." Id. "[T]rial judges may evaluate the data offered to support an expert's bottom-line opinions to determine if that data provides adequate support to mark the expert's testimony as reliable." Ruiz-Troche, 161 F.3d at 81.

This does not mean that trial courts are empowered "to determine which of several competing scientific theories has the best provenance." Id. at 85. "Daubert does not require that a party who proffers expert testimony carry the burden of proving to the judge that the expert's assessment of the situation is

correct." Id. The proponent of the evidence must show only that "the expert's conclusion has been arrived at in a scientifically sound and methodologically reliable fashion." Id.; see also United States v. Vargas, 471 F.3d 255, 265 (1st Cir. 2006). The object of Daubert is "to make certain that an expert, whether basing testimony on 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." Kumho Tire, 526 U.S. at 152.

So long as an expert's scientific testimony rests upon "'good grounds,' based on what is known," Daubert, 509 U.S. at 590, it should be tested by the adversarial process, rather than excluded for fear that jurors will not be able to handle the scientific complexities, id. at 596. "Vigorous cross-examination, presentation of contrary evidence, and careful instruction on the burden of proof are the traditional and appropriate means of attacking shaky but admissible evidence." Id.; see also Currier v. United Techs. Corp., 393 F.3d 246, 252 (1st Cir. 2004).

II.

It is uncontested that Dr. Smith's opinion about the causal link between benzene and APL satisfies certain requirements of Rule 702. His opinion would "assist the trier of fact to understand the evidence or to determine a fact in issue." Fed. R. Evid. 702. And Dr. Smith is "a witness qualified as an expert by

knowledge, skill, experience, training, or education." Id. He is acknowledged as a leading expert on the study of the toxic effects of chemicals and drugs on the human body, with particular emphasis on the mechanisms by which benzene and its metabolites cause damage to both cells and the human organism as a whole. The research in Dr. Smith's laboratory, which is funded by the National Institutes of Health, focuses on the causes of leukemia and lymphoma, and he has authored or co-authored over 215 articles in peer-reviewed journals in the field of toxicology.

The question before us is whether the district court abused its discretion in concluding that the other requirements of Rule 702, concerning the reliability of Dr. Smith's opinion, were not met. We will first discuss some basic facts about leukemia, the weight of the evidence methodology, and Dr. Smith's use of that methodology, and we will then turn to an evaluation of the district court's ruling.

Leukemia is a cancer of the blood cells. There are different types of leukemia, which are generally classified in two ways. The first classification is between leukemia's acute and chronic forms: acute leukemia is characterized by a rapid increase in the number of immature blood cells, while chronic leukemia is characterized by the excessive buildup of relatively mature but abnormal white blood cells. The second classification is between the types of stem cells affected: leukemia can be either "myeloid"

or "lymphoid." Combining these two classifications provides a total of four main categories of leukemia: acute myeloid leukemia (AML); chronic myeloid leukemia (CML); acute lymphoid leukemia (ALL); and chronic lymphoid leukemia (CLL).² Within each of these categories, there are typically several subcategories.

The general category of AML can be subdivided in more than one way. Under the common French-American-British classification system used by the parties, subtypes are classified morphologically according to the degree of differentiation along different cell lines and the extent of cell maturation. This classification system identifies subtypes by convention as M0 through M7.³

Brian Milward's leukemia, APL, is subtype M3 and is an extremely rare disease. APL accounts for only five to ten percent of all cases of AML, which is itself rare, with an annual incidence of 3.5 cases per 100,000 people. APL is characterized by a deficiency of mature blood cells in the myeloid cell line and an excess of immature cells called promyelocytes.

² There are also some types of leukemia that are considered to be outside of this four-part classification scheme.

³ The World Health Organization has adopted a different classification system that utilizes not only morphological characteristics, but also genetic, immunophenotypic, biologic, and clinical characteristics to define specific disease entities that have clinical and biological relevance. See generally James W. Vardiman et al., The World Health Organization (WHO) Classification of the Myeloid Neoplasms, 100 Blood 2292 (2002).

APL is in part caused by the chromosomal translocation of a gene known as the retinoic acid receptor-alpha gene (RAR α) on chromosome 17.⁴ Although APL and the other subtypes of AML have been the subject of extensive research, there is not yet a scientific consensus as to the causes of the genetic translocation that induces APL.

Dr. Smith's opinion is that what is known about both AML and APL supports the inference that exposure to benzene can cause APL. He reached this opinion using a "weight of the evidence" methodology in which he considered five lines of evidence drawn from the peer-reviewed scientific literature on leukemia and benzene. We first discuss the reliability of this methodology in general, and then turn to Dr. Smith's application of it.

A. The Reliability of the Weight of the Evidence Methodology

Dr. Smith's opinion was based on a "weight of the evidence" methodology in which he followed the guidelines articulated by world-renowned epidemiologist Sir Arthur Bradford Hill in his seminal methodological article on inferences of causality. See Arthur Bradford Hill, The Environment and Disease: Association or Causation?, 58 Proc. Royal Soc'y Med. 295 (1965).

⁴ In approximately 95% of cases of APL, RAR α is involved in a reciprocal translocation with the promyelocytic leukemia gene (PML) on chromosome 15--a translocation denoted as t(15;17)(q22;q12)--which creates a fusion gene known as PML-RAR α . In the remaining cases of APL, RAR α translocates and fuses with one of four other genes.

Hill's article explains that one should not conclude that an observed association between a disease and a feature of the environment (e.g., a chemical) is causal without first considering a variety of "viewpoints" on the issue. These viewpoints include: the strength or frequency of the association; the consistency of the association in varied circumstances; the specificity of the association; the temporal relationship between the disease and the posited cause; the dose response curve between them; the biological plausibility of the causal explanation given existing scientific knowledge; the coherence of the explanation with generally known facts about the disease; the experimental data that relates to it; and the existence of analogous causal relationships. See id. at 295-99.⁵

Although Hill identified nine viewpoints, it is generally agreed that this list is not exhaustive and that no one type of evidence must be present before causality may be inferred. For example, when a group from the National Cancer Institute was asked to rank the different types of evidence, it concluded that "[t]here

⁵ See also Sheldon Krimsky, The Weight of Scientific Evidence in Policy and Law, 95 Am. J. Pub. Health S129, S129 (2005) (explaining that the term "weight of the evidence" is used "to characterize a process or method in which all scientific evidence that is relevant to the status of a causal hypothesis is taken into account").

should be no such hierarchy."⁶ Michele Carbon et al., Modern Criteria to Establish Human Cancer Etiology, 64 Cancer Res. 5518, 5522 (2004); see also Sheldon Krinsky, The Weight of Scientific Evidence in Policy and Law, 95 Am. J. Pub. Health S129, S130 (2005).

This "weight of the evidence" approach to making causal determinations involves a mode of logical reasoning often described as "inference to the best explanation," in which the conclusion is not guaranteed by the premises.⁷ See Bitler v. A.O. Smith Corp., 391 F.3d 1114, 1124 n.5 (10th Cir. 2004). As explained by plaintiffs' expert on methodology Dr. Cranor, Distinguished

⁶ This point was also emphasized by Hill, who cautioned in his article:

None of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a sine qua non. What they can do, with greater or less strength, is to help us to make up our minds on the fundamental question--is there any other way of explaining the set of facts before us, is there any other answer equally, or more, likely than cause and effect?

Austin Bradford Hill, The Environment and Disease: Association or Causation?, 58 Proc. Royal Soc'y Med. 295, 299 (1965).

⁷ "Unlike a logical inference made by deduction where one proposition can be logically inferred from other known propositions, and unlike induction where a generalized conclusion can be inferred from a range of known particulars, inference to the best explanation--or 'abductive inferences'--are drawn about a particular proposition or event by a process of eliminating all other possible conclusions to arrive at the most likely one, the one that best explains the available data." Bitler v. A.O. Smith Corp., 391 F.3d 1114, 1124 n.5 (10th Cir. 2004).

Professor of Philosophy at the University of California, Riverside, inference to the best explanation can be thought of as involving six general steps, some of which may be implicit. The scientist must (1) identify an association between an exposure and a disease, (2) consider a range of plausible explanations for the association, (3) rank the rival explanations according to their plausibility, (4) seek additional evidence to separate the more plausible from the less plausible explanations, (5) consider all of the relevant available evidence, and (6) integrate the evidence using professional judgment to come to a conclusion about the best explanation.

In this mode of reasoning, the use of scientific judgment is necessary. "No algorithm exists for applying the Hill guidelines to determine whether an association truly reflects a causal relationship or is spurious." Restatement § 28 cmt. c(3). Because "[n]o scientific methodology exists for this process . . . reasonable scientists may come to different judgments about whether such an inference is appropriate." Id. § 28 reporters' note cmt. c(4).

The fact that the role of judgment in the weight of the evidence approach is more readily apparent than it is in other methodologies does not mean that the approach is any less scientific. No matter what methodology is used, "an evaluation of data and scientific evidence to determine whether an inference of

causation is appropriate requires judgment and interpretation." Id. § 28 cmt. c(1).⁸ The use of judgment in the weight of the evidence methodology is similar to that in differential diagnosis, see Cruz v. Bridgestone/Firestone N. Am. Tire, LLC, 388 F. App'x 803, 806-07 (10th Cir. 2010) (explaining that differential analysis in general is best characterized as a process of reasoning to the best explanation), which we have repeatedly found to be a reliable method of medical diagnosis, see Granfield v. CSX Transp., Inc., 597 F.3d 474, 486 (1st Cir. 2010); Dalkon Shield, 156 F.3d at 253.

Defendants argue that "regardless of its level of acceptance in the scientific community, a pure 'weight of the evidence' approach like that utilized by Dr. Smith . . . is hardly the type of reliable scientific evidence contemplated by Daubert."⁹

⁸ The fact that epidemiology relies on statistical methods does not avoid the use of judgment, as "[e]ven sampling error, which is analyzed using quantitative statistical methods, only provides a range of outcomes (associations) that might have been produced by sampling error even if there is no association between the agent and disease. Thus, interpreting the results of epidemiologic studies requires informed judgment and is subject to uncertainty." Restatement (Third) of Torts: Liability for Physical and Emotional Harm § 28 reporters' note cmt. c(3) (2010).

⁹ Defendants draw our attention to a Fifth Circuit case, excluding testimony based on a weight of the evidence methodology, in which the court explained:

We are also unpersuaded that the 'weight of the evidence' methodology these experts use is scientifically acceptable for demonstrating a medical link between Allen's EtO exposure and brain cancer. Regulatory and advisory bodies such as IARC, OSHA and EPA utilize a 'weight of the evidence' method to assess the carcinogenicity of various substances in human

No serious argument can be made that the weight of the evidence approach is inherently unreliable. Rather, admissibility must turn on the particular facts of the case. See, e.g., Cruz, 388 F. App'x at 807 (explaining that expert testimony based on "inference to the best explanation" may be admissible, but that there was no error in the district court's finding that the expert's specific theory did not have sufficient scientific support). Here, the question is whether Dr. Smith, in reaching his opinion, applied the methodology with "the same level of intellectual rigor" that he uses in his scientific practice. Kumho Tire, 526 U.S. at 152.

B. Dr. Smith's Application of the Methodology

In concluding that the weight of the evidence supported the conclusion that benzene can cause APL, Dr. Smith relied on his knowledge and experience in the field of toxicology and molecular epidemiology and considered five bodies of evidence drawn from the peer-reviewed scientific literature on benzene and leukemia.

beings and suggest or make prophylactic rules governing human exposure. This methodology results from the preventive perspective that the agencies adopt in order to reduce public exposure to harmful substances. The agencies' threshold of proof is reasonably lower than that appropriate in tort law

Allen v. Pa. Eng'g Corp., 102 F.3d 194, 198 (5th Cir. 1996). However, the Fifth Circuit did not, as defendants contend, hold "that the 'weight-of-the evidence' approach is per se unreliable." Rather, the court rejected its use in that case--a case in which it found that the experts' conclusion was "at best weakly supported, if not contradicted, by the evidence on which they rely," and in which the experts "all declined to say that they would subject their findings to the test of peer review for publication." Id.

First, Dr. Smith considered the near-consensus among governmental agencies, experts, and active researchers in the field that benzene can cause AML as a class. The existence of this causal connection has been established since the late 1970s. See Bernard D. Goldstein & Gisela Witz, Benzene, in Environmental Toxicants: Human Exposures and Their Health Effects 459, 478 (Morton Lippmann ed., 3rd ed. 2009). Dr. Smith noted that epidemiological studies have found a statistically significant increased incidence of AML in benzene-exposed workers and have identified a dose-response relationship.

Second, Dr. Smith considered evidence concerning the etiology, or origins, of leukemia indicating that all types of AML derive from a genetically damaged pluripotent stem cell. Dr. Smith referred to a recent peer-reviewed article that provided a review of the current literature and reported numerous studies demonstrating that both AML and CML are stem cell diseases. He cited peer-reviewed studies finding that in the APL and Core Binding Factor (CBF) subtypes of AML, as well as in CML, the stem cell mutation is often in part caused by a chromosomal translocation. He also cited evidence that APL and CBF share common genetic susceptibility factors, common risk factors, and the same incidence pattern occurring at a constant incidence with age after age 20. Dr. Smith concluded that the best explanation for

this evidence is that all AMLs, including APL, have a common etiology.¹⁰

Third, Dr. Smith considered toxicology studies establishing that metabolites of benzene cause significant chromosomal damage at the stem cell level in the bone marrow--the type of damage that is known to cause APL and other types of AML.¹¹ He also cited peer-reviewed work published by his lab showing that leukemia cases associated with benzene exposure are more likely to contain clonal chromosome aberrations than leukemias arising in the general population.

Fourth, Dr. Smith considered two sets of studies concerning the inhibition of a cellular enzyme known as topoisomerase II (or "topo II") that is essential for the maintenance of proper chromosome structure and segregation. One set of studies--including both test tube and animal studies--has established that two benzene metabolites are catalytic inhibitors

¹⁰ Defendants' experts questioned Dr. Smith's conclusion that all of the subtypes of AML have a common etiology. However, on cross examination in the district court Daubert hearing, defendants' expert Dr. Pyatt agreed with the statement that "there are a group of reasonable scientists who reasonably believe that all forms of AML arise from the same progenitor cell" and stated that Dr. Smith's opinion was "consistent with most of the evidence." Defendants' expert Dr. Bennett likewise agreed that "reasonable scientists can and do" agree with Dr. Smith.

¹¹ Defendants' expert Dr. Bennett agreed that "there have been innumerable studies that demonstrate that benzene actually works at multiple levels to create damage to the DNA structure of this hematopoietic stem cell."

of topo II. A second set of studies has established that a variety of chemotherapeutic agents that are catalytic inhibitors of topo II cause APL.¹² Dr. Smith explained that taken together, these studies provided evidence of a known biological mechanism by which exposure to benzene could cause APL.

Fifth, Dr. Smith considered the small set of epidemiological studies that provide data on the relationship between benzene exposure and subtypes of AML.¹³ He concluded that the evidence showed an increased risk factor for APL, consistent with causality, and provided no grounds for concluding otherwise.

Dr. Smith explained that taking into account all of the evidence described above--the fact that benzene causes AML as a class, that all subtypes of AML likely have a common etiology, that

¹² Dr. Smith cited a long list of peer-reviewed publications and quoted a recent authoritative paper in a prominent journal stating that "[t]herapy-related acute promyelocytic leukemia (t-APL) with the t(15;17) translocation is a well recognized complication of cancer treatment with agents targeting topoisomerase II." Syed Khizer Hasan et al., Molecular Analysis of t(15;17) Genomic Breakpoints in Secondary Acute Promyelocytic Leukemia Arising After Treatment of Multiple Sclerosis, 112 Blood 3383, 3383 (2008). Defendants' hematopathologist, Dr. Bennett, acknowledged that chemotherapeutic compounds that inhibit topo II can cause APL.

¹³ He considered a multi-center Chinese case-control study of 1257 cases of leukemia in which there was a 40% increased risk of APL in benzene-exposed workers; a cohort study of 74,828 workers exposed to benzene in China in which APL was the most common form of AML diagnosed; a multi-center Italian case-control study of 38 cases of APL that showed a strong association between APL and shoe-making, an industry that had for many years used benzene as an adhesive; and several case reports of APL in benzene-exposed workers.

benzene is known to cause the type of chromosomal damage characteristic of APL, that benzene is known to inhibit an enzyme whose inhibition is known to cause APL, and that APL has been reported in benzene-exposed workers in a number of epidemiological studies--he reached the opinion that the weight of the evidence supports the conclusion that benzene exposure is capable of causing APL. Dr. Smith's opinion rests on a scientifically sound and methodologically reliable foundation, as is required by Daubert.

III.

In finding Dr. Smith's opinion inadmissible under Rule 702, the district court relied on (a) its evaluation of the mechanistic and epidemiological evidence on which Dr. Smith based his opinion, and (b) its understanding of the scientific concept of "biological plausibility" as used by Dr. Smith when he explained his conclusions. As we explain below, on both of these points, the district court erred. In the end, the court's exclusion of the testimony was based on its evaluation of the weight of the evidence, which is an issue that is the province of the jury, and on its misperception of the methodology and analysis that provided the basis for Dr. Smith's opinion.

A. The Evidentiary Basis of Dr. Smith's Opinion

1. Mechanistic Evidence

The district court's exclusion of Dr. Smith's testimony was based to a significant extent on its rejection of what it took

to be his three key subsidiary conclusions regarding the weight of the mechanistic evidence. We briefly summarize the court's analysis on these points before turning to our discussion of the ways in which the court erred in its analysis.

First, the court held that there was insufficient evidence to support Dr. Smith's opinion that all subtypes of AML likely have a common etiology. The court reasoned that the "clear differences" among AML subtypes--in particular, APL's unique response to certain types of therapy, and the subtypes' different chromosomal abnormalities--made "a broad extrapolation from AML generally to APL specifically" inappropriate.¹⁴ Milward, 664 F. Supp. 2d at 144. The court also noted that a series of recent studies had "led investigators to think that the 'leukemic stem cell' may exist in more mature, differentiated cell lines," such that "the 'leukemic stem cell' may not be a stem cell in the usual sense, but rather a differentiated cell that has somehow acquired the ability to reproduce itself, as a stem cell can." Id. at 145. If the various AML subtypes did not arise from the same progenitor or stem cell, the court reasoned, they might well not share a

¹⁴ Dr. Smith's supplemental report makes it clear that this is something on which reasonable scientific disagreement is possible. He explained that in his view, defendants' experts erred in concluding that the fact that APL is therapeutically unique means that it is also etiologically unique. Identifying the biological mechanism that made APL therapeutically unique--the sensitivity of the PML-RAR α fusion gene to retinoic acid and arsenic--he explained that this was "irrelevant" to APL's etiology.

common etiology.¹⁵ Finally, the court emphasized that there was "no scientific consensus" on this issue, and that the question of when the key chromosomal translocation occurs was considered by researchers to be "a question that remains unanswered in the APL field." Id. (quoting S. Wojiski et al., PML-RAR α Initiates Leukemia by Conferring Properties of Self-Renewal to Committed Promyelocytic Progenitors, 23 *Leukemia* 1462, 1469 (2009) (emphasis added)) (internal quotation marks omitted).

Second, the court held that what was known about the types of chromosomal translocations caused by benzene did not offer sufficient support for Dr. Smith's opinion that it is biologically plausible that benzene causes the characteristic t(15;17)(q22;q12) translocation seen with APL. The court explained that this opinion would be warranted if benzene's impact on chromosomes were randomly experienced, but it noted that a paper co-authored by Dr. Smith concluded that "benzene can initiate or promote leukemia induction by a nonrandom selective effect" on specific chromosomes. Id. at 147 (emphasis added). This defeated "the generalization that because . . . benzene causes damage to some chromosomes, it is

¹⁵ In the Daubert hearing, Dr. Smith made it clear that he had considered the key paper cited by the district court on this point. He noted that it was based on studies in "mice using a highly artificial system," and he explained that even if the mutation could occur at a later point in differentiation as indicated by this paper, "it doesn't mean that it has to occur only in that compartment."

'biologically plausible' that it causes damage to other chromosomes." Id.

Third, the court held that there was insufficient evidence to support the inference that benzene metabolites inhibit topo II in such a way as to cause the chromosomal translocation seen in cases of APL. The court's conclusion was in part based on evidence that "[t]here are different classes of topo II inhibitors and the different classes have been associated with different AML subtypes." Id. Highlighting one article's finding that leukemias induced by benzene do not appear to exhibit the defining characteristics associated with four other classes of topo II inhibitors, id. at 148, the court held that to "the extent that Dr. Smith's opinion rests on the proposition that all topo II inhibitors act similarly to cause a similar effect, then, it does not appear to be based on reliable scientific knowledge," id. at 147.

In reaching these three conclusions about some of the evidence on which Dr. Smith based his opinion, the court both placed undue weight on the lack of general acceptance of Dr. Smith's conclusions and crossed the boundary between gatekeeper and trier of fact.

Although general acceptance is still a relevant consideration under Daubert, the court's demands went too far. Cf. Smith v. Ford Motor Co., 215 F.3d 713, 721 (7th Cir. 2000)

(reversing district court that had treated lack of peer review as dispositive grounds for excluding expert opinion). On the question of the origins of APL, for example, the court explained that in the absence of consensus about the target cell for the leukemic mutation, Dr. Smith's opinion that all forms of AML likely share a common origin was "at best a plausible hypothesis." Milward, 664 F. Supp. 2d at 146. The court explained that the fact that "other plausible hypotheses . . . might be true as well, including the hypothesis that the genetic mutation that leads to APL occurs in relatively mature cells," meant that Dr. Smith's opinion was not "based on sufficient facts and data to be accepted as a reliable scientific conclusion." Id.; see also id. at 148 (focusing on lack of consensus as to the topo II question). But the fact that another explanation might be right is not a sufficient basis for excluding Dr. Smith's testimony. "Lack of certainty is not, for a qualified expert, the same thing as guesswork." Primiano v. Cook, No. 06-15563, 2010 WL 1660303, *5 (9th Cir. Apr. 27, 2010).

In addition, the alleged flaws identified by the court go to the weight of Dr. Smith's opinion, not its admissibility. There is an important difference between what is unreliable support and what a trier of fact may conclude is insufficient support for an expert's conclusion.

The court's analysis repeatedly challenged the factual underpinnings of Dr. Smith's opinion, and took sides on questions

that are currently the focus of extensive scientific research and debate--and on which reasonable scientists can clearly disagree. In this, the court overstepped the authorized bounds of its role as gatekeeper. "The soundness of the factual underpinnings of the expert's analysis and the correctness of the expert's conclusions based on that analysis are factual matters to be determined by the trier of fact." Smith, 215 F.3d at 718. "When the factual underpinning of an expert's opinion is weak, it is a matter affecting the weight and credibility of the testimony--a question to be resolved by the jury." Vargas, 471 F.3d at 264 (quoting Int'l Adhesive Coating Co. v. Bolton Emerson Int'l, 851 F.2d 540, 545 (1st Cir. 1988)) (internal quotation marks omitted); see also Quiet Tech. DC-8, Inc. v. Hurel-Dubois UK Ltd., 326 F.3d 1333, 1345 (11th Cir. 2003); Amorgianos v. Nat'l R.R. Passenger Corp., 303 F.3d 256, 267 (2d Cir. 2002).

Of course, following Joiner, a "district court properly may exclude expert testimony if the court concludes too great an analytical gap exists between the existing data and the expert's conclusion." Kennedy v. Collagen Corp., 161 F.3d 1226, 1230 (9th Cir. 1998). Here, however, "the gap was of the district court's making." Id. Dr. Smith's opinion was based on a reliable methodology and substantial evidence that he carefully explained. The questions that the court posed were sensible ones, but ones for the jury to resolve.

_____ At times, the court's error in excluding Dr. Smith's testimony derived from a mistake in its understanding of the weight of the evidence methodology employed by Dr. Smith. The court treated the separate evidentiary components of Dr. Smith's analysis atomistically, as though his ultimate opinion was independently supported by each. For example, the court referred to "Dr. Smith's opinion that because benzene metabolites inhibit topo II and because some classes of topo II inhibitors appear to have a causal relationship to APL, therefore benzene has a causal relationship to APL." Milward, 664 F. Supp. 2d at 148 (emphasis added). This overstates Dr. Smith's conclusion as to the topo II evidence, and is indicative of an error in the court's understanding of the nature of Dr. Smith's analysis.

In Dr. Smith's weight of the evidence approach, no body of evidence was itself treated as justifying an inference of causation. Rather, each body of evidence was treated as grounds for the subsidiary conclusion that it would, if combined with other evidence, support a causal inference. The district court erred in reasoning that because no one line of evidence supported a reliable inference of causation, an inference of causation based on the totality of the evidence was unreliable. Cf. NutraSweet Co. v. X-L Eng'g Co., 227 F.3d 776, 789 (7th Cir. 2000) (holding that an expert's reliance on individual pieces of evidence, insufficient in themselves to prove a point, "did not render his opinion

speculative").¹⁶ The hallmark of the weight of the evidence approach is reasoning to the best explanation for all of the available evidence. Cf. Dalkon Shield, 156 F.3d at 253 (reversing district court's exclusion of expert testimony as "guesswork" or without "basis" when testimony was based on differential diagnosis and there was no showing that any one of the expert's premises was "so faulty that it could not even be tendered to the jury for its consideration"); see also Hardyman v. Norfolk & W. Ry. Co., 243 F.3d 255, 261 (6th Cir. 2001).

2. Epidemiological Evidence

As to the epidemiological evidence on which Dr. Smith based his opinion in part, the court held that the published articles on which Dr. Smith relied did not support his opinion, and that in any event, the evidence was not statistically significant. On these grounds, the court rejected Dr. Smith's conclusion that the available epidemiological evidence offered some support for an inference of causation.

In concluding that the papers cited by Dr. Smith did not support his opinion, the court reasoned that "Dr. Garabrant convincingly demonstrated, especially with respect to the Golomb

¹⁶ As a general evidentiary matter, "individual pieces of evidence, insufficient in themselves to prove a point, may in cumulation prove it," and "a piece of evidence, unreliable in isolation, may become quite probative when corroborated by other evidence." Bourjaily v. United States, 483 U.S. 171, 179-80 (1987).

and Travis papers, that Dr. Smith's conclusions that there was a positive association between exposure to benzene and APL were based on faulty calculations of odds ratios." Milward, 664 F. Supp. 2d at 149. An odds ratio represents the difference in the incidence of a disease between a population that has been exposed to benzene and one that has not. In Dr. Garabrant's opinion, Dr. Smith should have used the incidence rate of APL for the general population as a baseline, rather than the rate for non-benzene-exposed workers. In the Daubert hearing and in his supplemental report, however, Dr. Smith explained that he disagreed with Dr. Garabrant on this point, but that in any event, the odds ratio was still elevated, consistent with an inference of causation. Where, as here, both experts' opinions are supported by evidence and sound scientific reasoning, the question of who is right is a question for the jury.¹⁷

The court explained, however, that even if "some of the data reported in the various studies could be properly understood to suggest a positive association, the findings are not statistically significant," id., and that although "epidemiological

¹⁷ The court also rejected Dr. Smith's analysis of the epidemiological evidence on the grounds that "none of the studies purports to give direct support to the proposition that benzene causes APL." Milward v. Acuity Specialty Prods. Grp., Inc., 664 F. Supp. 2d 137, 148 (D. Mass. 2009). Yet Dr. Smith did not claim that the studies provided direct support. Rather, his characterization of his methodology makes clear that he was using them as indirect support.

evidence is not always essential," the defendants were "correct that sound epidemiological studies are ordinarily needed to confirm, by consistent observation, an hypothesis of causation," id. at 148.

In context, the district court read too much into the paucity of statistically significant epidemiological studies. The absence of peer-reviewed epidemiological studies does not, as defendants contend, make it "almost impossible" for Dr. Smith's opinion to be admissible. Epidemiological studies are not per se required as a condition of admissibility regardless of context. See Rider v. Sandoz Pharm. Corp., 295 F.3d 1194, 1198 (11th Cir. 2002) ("It is well-settled that while epidemiological studies may be powerful evidence of causation, the lack thereof is not fatal to a plaintiff's case."); Restatement § 28 reporters' note cmt. c(3) (listing federal circuit cases holding that epidemiological data is not necessary). Nor are such studies treated as always essential in the relevant scientific communities.

To be clear, this is not a situation in which the available epidemiological studies found that there is no causal link, or even one in which no cases of APL were found among benzene-exposed workers. Cf. Norris v. Baxter Healthcare Corp., 397 F.3d 878, 882 (10th Cir. 2005) (holding that epidemiological studies are not required to prove causation, but that a substantial body of epidemiological evidence challenging causation cannot be

ignored); Allen v. Pa. Eng'g Corp., 102 F.3d 194, 197 (5th Cir. 1996) (finding it significant that "numerous reputable epidemiological studies covering in total thousands of workers" indicated that there was no causation).

Rather, this is a case in which the few studies that differentiate between AML and APL do not offer conclusive statistically significant evidence either way, in part because the rarity of APL makes it nearly impossible to perform a large enough study.¹⁸ Dr. Smith estimated that in order to obtain statistically significant results, one would need hundreds of thousands of highly exposed workers, the same number of controls, and millions of dollars in funding. The court erred in treating the lack of statistical significance as a crucial flaw. See Collagen Corp., 161 F.3d at 1229 (finding that the district court placed too much emphasis on lack of epidemiological studies where such studies "would be almost impossible to perform"); see also Primiano, 2010 WL 1660303, at *5-6 (noting that peer-reviewed studies are not necessary, especially when there are good reasons why such studies have not been performed). Under these circumstances, the court

¹⁸ The difficulty of performing such a study is not contested by defendants, and it has even been expressly affirmed in the scientific literature. See Dan Douer, The Epidemiology of Acute Promyelocytic Leukaemia, 16 Best Prac. & Res. Clinical Haematology 357, 358 (2003) ("It is difficult to perform epidemiological studies in AML subtypes classified according to cytogenetic abnormalities owing to the small number of patients within each subgroup.").

erred in holding that "Dr. Smith's attempt to support his conclusion with data that concededly lacks statistical significance" was "a deviation from sound practice of the scientific method" that provided grounds for exclusion. Milward, 664 F. Supp. 2d at 149.

The court's evaluation of the epidemiological evidence is also in tension with the weight of the evidence methodology. Dr. Smith explained that his citation to epidemiological data was meant to challenge the theory that benzene exposure could not cause APL, and to highlight that the limited data available was consistent with the conclusions that he had reached on the basis of other bodies of evidence. He stated that "[i]f epidemiologic studies of benzene-exposed workers were devoid of workers who developed APL, one could hypothesize that benzene does not cause this particular subtype of AML." The fact that, on the contrary, "APL is seen in studies of workers exposed to benzene where the subtypes of AML have been separately analyzed and has been found at higher levels than expected" suggested to him that the limited epidemiological evidence was at the very least consistent with, and suggestive of, the conclusion that benzene can cause APL.

The court rejected Dr. Smith's reasoning, stating that a "'suggestion' may give rise to a plausible hypothesis, but not a reliable inference." Milward, 664 F. Supp. 2d at 149. But as noted above, this is inconsistent with the scientifically accepted

methodology employed by Dr. Smith. Dr. Smith did not infer causality from this suggestion alone, but rather from the accumulation of multiple scientifically acceptable inferences from different bodies of evidence.

B. The Concept of "Biological Plausibility"

The district court also erred in its apprehension of the scientific concept of biological plausibility and its place in Dr. Smith's analysis. The concept of biological plausibility, which numbers among the nine Hill viewpoints, asks whether the hypothesized causal link is credible in light of what is known from science and medicine about the human body and the potentially offending agent. At two places in the court's analysis, it conflated the scientific question of biological plausibility with the legal question of probability.

In the court's discussion of the epidemiological evidence, it stated that even if the evidence "'suggests' a causal relationship," providing support for Dr. Smith's opinion regarding biological plausibility, a "plausible hypothesis" is not a "reliable inference" and is therefore inadmissible. Id. Here, the court not only misconstrued the concept of biological plausibility by equating it with a merely plausible or possible hypothesis, but also misconstrued the concept's role in Dr. Smith's analysis by assuming that Dr. Smith treated the criteria as sufficient grounds

for inferring causality (rather than as one consideration that entered into his weighing of the evidence).

The court made a similar error in its conclusion, where it stated:

While Dr. Smith's hypotheses are, to use his term, "plausible," they remain hypotheses, the validity of which has not been reliably established. . . . [T]he sum of Dr. Smith's testimony, fairly understood, is that benzene might be a cause of APL.

Id. Again, the district court misunderstood Dr. Smith to be saying that causation is possible rather than probable. The sum of Dr. Smith's testimony was not merely that it is possible, or even biologically plausible, that benzene causes APL. Rather, the sum of his testimony was that a weighing of the Hill factors, including biological plausibility, supported the inference that the association between benzene exposure and APL is genuine and causal.

IV.

The record clearly demonstrates that Dr. Smith's opinion was based on an analysis in which he employed the "same level of intellectual rigor" that he employs in his academic work. Kumho Tire, 526 U.S. at 152. In excluding Dr. Smith's testimony, the district court did not properly apply Daubert and exceeded the scope of its discretion. We reverse the district court's judgment

for the defendants and its exclusion of Dr. Smith's testimony, and we remand for proceedings consistent with this opinion.¹⁹

So ordered.

¹⁹ We wish to acknowledge the able briefing of the issues by the parties and amici.