

PUBLISHED

UNITED STATES COURT OF APPEALS
FOR THE FOURTH CIRCUIT

No. 17-1140

In Re: LIPITOR (ATORVASTATIN CALCIUM) MARKETING,
SALES PRACTICES AND PRODUCTS LIABILITY LITIGATION
(NO II) MDL 2502,

PLAINTIFFS APPEALING CASE MANAGEMENT ORDER 100,

Plaintiffs – Appellants,

v.

PFIZER, INCORPORATED; MCKESSON CORPORATION; GREENSTONE,
LLC; PFIZER INTERNATIONAL LLC,

Defendants – Appellees,

JONAH B. GELBACH; CARL CRANOR; DIERDRE N. MCCLOSKEY;
STEPHEN T. ZILIAK,

Amici Supporting Appellants,

PRODUCT LIABILITY ADVISORY COUNCIL, INCORPORATED;
WASHINGTON LEGAL FOUNDATION; CHAMBER OF COMMERCE OF
THE UNITED STATES OF AMERICA; PHARMACEUTICAL RESEARCH
AND MANUFACTURERS OF AMERICA; AMERICAN TORT REFORM
ASSOCIATION,

Amici Supporting Appellees.

No. 17-1136

In Re: LIPITOR (ATORVASTATIN CALCIUM) MARKETING,
SALES PRACTICES AND PRODUCTS LIABILITY LITIGATION
(NO II) MDL 2502,

JUANITA HEMPSTEAD,

Plaintiff – Appellant,

v.

PFIZER, INCORPORATED; PFIZER INTERNATIONAL LLC; GREENSTONE,
LLC,

Defendants – Appellees,

JONAH B. GELBACH; CARL CRANOR; DIERDRE N. MCCLOSKEY;
STEPHEN T. ZILIAK,

Amici Supporting Appellants,

PRODUCT LIABILITY ADVISORY COUNCIL, INCORPORATED;
WASHINGTON LEGAL FOUNDATION; CHAMBER OF COMMERCE OF
THE UNITED STATES OF AMERICA; PHARMACEUTICAL RESEARCH
AND MANUFACTURERS OF AMERICA; AMERICAN TORT REFORM
ASSOCIATION,

Amici Supporting Appellees.

No. 17-1137

In Re: LIPITOR (ATORVASTATIN CALCIUM) MARKETING, SALES
PRACTICES AND PRODUCTS LIABILITY LITIGATION (NO II)
MDL 2502,

PLAINTIFFS APPEALING CASE MANAGEMENT ORDER 99,

Plaintiffs – Appellants,

v.

PFIZER, INCORPORATED; MCKESSON CORPORATION; GREENSTONE, LLC; PFIZER INTERNATIONAL LLC,

Defendants – Appellees,

JONAH B. GELBACH; CARL CRANOR; DIERDRE N. MCCLOSKEY; STEPHEN T. ZILIAK,

Amici Supporting Appellants,

PRODUCT LIABILITY ADVISORY COUNCIL, INCORPORATED; WASHINGTON LEGAL FOUNDATION; CHAMBER OF COMMERCE OF THE UNITED STATES OF AMERICA; PHARMACEUTICAL RESEARCH AND MANUFACTURERS OF AMERICA; AMERICAN TORT REFORM ASSOCIATION,

Amici Supporting Appellees.

No. 17-1189

In Re: LIPITOR (ATORVASTATIN CALCIUM) MARKETING, SALES PRACTICES AND PRODUCTS LIABILITY LITIGATION (NO II) MDL 2502,

PLAINTIFFS APPEALING CASE MANAGEMENT ORDER 109,

Plaintiffs – Appellants,

v.

PFIZER, INCORPORATED; GREENSTONE, LLC; PFIZER INTERNATIONAL LLC; MCKESSON CORPORATION,

Defendants – Appellees,

JONAH B. GELBACH; CARL CRANOR; DIERDRE N. MCCLOSKEY,
STEPHEN T. ZILIAK,

Amici Supporting Appellants,

PRODUCT LIABILITY ADVISORY COUNCIL, INCORPORATED;
WASHINGTON LEGAL FOUNDATION; CHAMBER OF COMMERCE OF
THE UNITED STATES OF AMERICA; PHARMACEUTICAL RESEARCH
AND MANUFACTURERS OF AMERICA; AMERICAN TORT REFORM
ASSOCIATION,

Amici Supporting Appellees.

Appeal from the United States District Court for the District of South Carolina, at
Charleston. Richard Mark Gergel, District Judge. (2:14-mn-02502-RMG)

Argued: January 23, 2018

Decided: June 12, 2018

Before NIEMEYER, KING, and DIAZ, Circuit Judges.

Affirmed by published opinion. Judge Diaz wrote the opinion, in which Judge Niemeyer
and Judge King joined.

ARGUED: Derek T. Ho, KELLOGG, HANSEN, TODD, FIGEL & FREDERICK,
P.L.L.C., Washington, D.C., for Appellants. Mark Cheffo, QUINN EMANUEL
URQUHART & SULLIVAN LLP, New York, New York, for Appellees. **ON BRIEF:** H.
Blair Hahn, Christiaan A. Marcum, RICHARDSON, PATRICK, WESTBROOK &
BRICKMAN, LLC, Mt. Pleasant, South Carolina; Silvija A. Strikis, Hilary P. Gerzhoy,
KELLOGG, HANSEN, TODD, FIGEL & FREDERICK, P.L.L.C., Washington, D.C., for
Appellants. Sheila L. Birnbaum, Bert L. Wolff, Mara Cusker Gonzalez, Lincoln Davis
Wilson, Jonathan S. Tam, QUINN EMANUEL URQUHART & SULLIVAN LLP, New
York, New York; Michael T. Cole, Charleston, South Carolina, David E. Dukes, NELSON
MULLINS RILEY & SCARBOROUGH LLP, Columbia, South Carolina, for Appellees
Pfizer Incorporated, Pfizer International LLC, and Greenstone LLC. Habib Nasrullah,
WHEELER TRIGG O'DONNELL LLP, Denver, Colorado, for Appellee McKesson
Corporation. Matthew Duncan, FINE, KAPLAN AND BLACK, R.P.C., Philadelphia,
Pennsylvania, for Amicus Jonah B. Gelbach. Christopher J. McDonald, Christopher D.
Barraza, LABATON SUCHAROW LLP, New York, New York, for Amici Carl Cranor,

Dierdre N. McCloskey, and Stephen T. Ziliak. Mary A. Wells, L. Michael Brooks, Jr., Brendan L. Loy, WELLS, ANDERSON & RACE, LLC, Denver, Colorado; Terri S. Reiskin, DYKEMA GOSSETT PLLC, Washington, D.C.; Hugh F. Young, Jr., PRODUCT LIABILITY ADVISORY COUNCIL, INC., Reston, Virginia, for Amicus Product Liability Advisory Council, Incorporated. Warren Postman, UNITED STATES CHAMBER LITIGATION CENTER, Washington, D.C.; Brian D. Boone, Emily C. McGowan, Charlotte, North Carolina, David Venderbush, ALSTON & BIRD LLP, New York, New York, for Amicus The Chamber of Commerce. Cory L. Andrews, Richard A. Samp, Mark S. Chenoweth, WASHINGTON LEGAL FOUNDATION, Washington, D.C., for Amicus Washington Legal Foundation. Eric G. Lasker, Kirby T. Griffis, Gregory S. Chernack, HOLLINGSWORTH LLP, Washington, D.C., for Amici American Tort Reform Association and Pharmaceutical Research and Manufacturers of America.

DIAZ, Circuit Judge:

This appeal arises from a multidistrict litigation (“MDL”) in which thousands of women claim that their use of the medication Lipitor caused them to develop diabetes. The women sued Pfizer, Lipitor’s manufacturer, asserting various products liability claims. After protracted litigation, the district court granted summary judgment to Pfizer. Plaintiffs now ask us to consider a host of issues, including whether the district court erred in excluding certain expert testimony under Federal Rule of Evidence 702 and *Daubert*¹; whether it erred in requiring expert testimony at all; and whether summary judgment was appropriately granted against all plaintiffs in the MDL. Finding no reversible error, we affirm the district court’s judgments.

I.

Pfizer manufactures Lipitor (atorvastatin calcium), a pharmaceutical drug. Lipitor is a member of a class of drugs known as statins, which are broadly indicated to prevent the onset of cardiovascular disease. Physicians prescribe Lipitor to lower patients’ low-density lipoprotein cholesterol (LDL-C, or “bad” cholesterol) and triglycerides in order to reduce the risk of heart attack or stroke. In the United States, Lipitor is commercially available in 10, 20, 40, and 80 mg tablets.

The plaintiffs in this litigation are more than three thousand women who have sued Pfizer, claiming that they developed diabetes as a result of taking Lipitor. Their complaint

¹ *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579 (1993).

sets forth various theories of liability, including that Pfizer was negligent in its design and promotion of Lipitor and that Pfizer failed to adequately warn others of the drug's known risks.

The Judicial Panel on Multidistrict Litigation transferred these lawsuits to the District of South Carolina for consolidated or coordinated pretrial proceedings. *See* 28 U.S.C. § 1407. The district court and the parties then agreed on four plaintiffs whose claims would serve as bellwether cases.

The parties engaged in extensive discovery, including the identification of expert witnesses and exchange of expert reports. The plaintiffs enlisted general causation experts, who intended to testify that there was a causal association between Lipitor and diabetes, and specific causation experts, who would testify that Lipitor proximately caused the onset of diabetes in each of the bellwether plaintiffs. The plaintiffs also retained an expert biostatistician, who performed analyses of several clinical trials and studies, and concluded that Lipitor led to a statistically significant increased risk of diabetes among those who took the drug.

The plaintiffs offered other evidence to prove causation. Specifically, they sought to introduce internal Pfizer emails, information from the Lipitor labeling in the United States and Japan, a statement in the New Drug Application (“NDA”) for Lipitor submitted by its original developer to the Food and Drug Administration (“FDA”), and information contained on the official Lipitor website. All of these, the plaintiffs contend, evince an association between Lipitor and diabetes—and Pfizer's knowledge of it.

At the close of discovery, Pfizer moved to exclude the plaintiffs' expert witnesses under *Daubert* and Federal Rule of Evidence 702. Following extensive hearings and an opportunity for the experts to amend their reports, Pfizer's challenge, in large part, succeeded. Relevant to this appeal, the district court excluded the opinions of the plaintiffs' statistician, Dr. Nicholas Jewell; the opinions of their general causation expert, Dr. Sonal Singh, except for his opinions relating to the 80 mg dose of Lipitor; and the specific causation opinions of Dr. Elizabeth Murphy, which related to the onset of diabetes in one of the bellwether plaintiffs.²

The court's rulings left the plaintiffs without their bellwether cases, and limited to a subset of patients who had taken an 80 mg dose. Following a hearing, and with agreement of counsel, the district court issued a series of four show cause orders asking whether any plaintiff in the MDL could submit evidence (expert or otherwise) that would enable her claim to survive summary judgment given the court's prior rulings.

In response to the show cause orders, one group of plaintiffs submitted evidence showing only that they were not diabetic before taking Lipitor, that they were diagnosed with diabetes after taking Lipitor, and that they lacked certain risk factors that might make them especially likely to develop the disease. Another group simply "dumped boxes upon boxes of documents" on the district court, including wholly irrelevant records such as "pictures from colonoscopies, EKGs, and pap smear results," with "no discernment or

² The district court struck additional general and specific causation experts, but the plaintiffs do not challenge those rulings.

suggestion as to which documents they claimed precluded summary judgment.” *In re Lipitor Mktg., Sales Practices and Prods. Liab. Litig.*, 226 F. Supp. 3d 557, 566 (D.S.C. 2017) (“CMO 99”).³ The district court determined that neither of these submissions were sufficient to show causation. *See id.* at 582.

After the court’s deadline to submit new evidence expired, the plaintiffs argued that the cases in the MDL ought to be returned to their transferor district courts for individual resolution on the issue of specific causation. The district court, however, deemed itself competent to address the questions that remained and granted summary judgment against all plaintiffs in the MDL.

II.

The lion’s share of this appeal centers upon the district court’s decision to exclude the testimony of three of the plaintiffs’ expert witnesses.

Expert testimony in the federal courts is governed by Federal Rule of Evidence 702. That rule permits an expert to testify where the expert’s “scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue,” so long as the expert’s opinion is “based on sufficient facts or data,” “is the product of reliable principles and methods,” and the expert “has reliably applied the principles and methods to the facts of the case.” Fed. R. Evid. 702.

³ For efficiency and clarity, we will use throughout our opinion the abbreviation “CMO” followed by the order number to differentiate among the many case management orders issued by the district court in this litigation.

In assessing the admissibility of expert testimony, a district court assumes a “gatekeeping role” to ensure that the “testimony both rests on a reliable foundation and is relevant to the task at hand.” *Daubert*, 509 U.S. at 597. The district court’s inquiry is a “flexible one,” whose focus “must be solely on principles and methodology, not on the conclusions that they generate.” *Id.* at 594–95. *Daubert*’s design is to “make certain that an expert, whether basing testimony upon professional studies or personal experience, employs in the courtroom the same level of intellectual rigor that characterizes the practice of an expert in the relevant field.” *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 152 (1999). A district court may consider a wide variety of factors to evaluate the reliability of expert testimony, including “testing, peer review, error rates, and ‘acceptability’ in the relevant scientific community.” *Id.* at 141.

“[T]he trial court’s role as a gatekeeper is not intended to serve as a replacement for the adversary system, and consequently, the rejection of expert testimony is the exception rather than the rule.” *United States v. Stanley*, 533 F. App’x 325, 327 (4th Cir. 2013) (per curiam) (unpublished) (internal quotation marks omitted). Indeed, *Daubert* itself stressed the importance of the “conventional devices” of “[v]igorous cross-examination, presentation of contrary evidence, and careful instruction on the burden of proof” (rather than wholesale exclusion by the trial judge) as “the traditional and appropriate means of attacking shaky but admissible evidence.” 509 U.S. at 596; *see also id.* at 601 (Rehnquist, C.J., concurring in part and dissenting in part) (cautioning that Rule 702 does not impose upon district judges “the obligation or the authority to become amateur scientists”).

When an expert brings science from the laboratory to the courthouse, we have recognized “two guiding, and sometimes competing, principles” that apply. *Westberry v. Gislaved Gummi AB*, 178 F.3d 257, 261 (4th Cir. 1999). “On the one hand . . . Rule 702 was intended to liberalize the introduction of relevant expert evidence. And, the court need not determine that the expert testimony a litigant seeks to offer into evidence is irrefutable or certainly correct.” *Id.* (citation omitted). But courts must also recognize that “due to the difficulty of evaluating their testimony, expert witnesses have the potential to be both powerful and quite misleading. And, given the potential persuasiveness of expert testimony, proffered evidence that has a greater potential to mislead than to enlighten should be excluded.” *Id.* (citation and internal quotation marks omitted).

We review a district court’s decision to admit or exclude expert evidence for abuse of discretion. *Gen. Elec. Co. v. Joiner*, 522 U.S. 136, 146 (1997). “If the district court makes an error of law in deciding an evidentiary question, that error is by definition an abuse of discretion. A district court likewise abuses its discretion in deciding a *Daubert* challenge if its conclusion rests upon a clearly erroneous factual finding.” *Nease v. Ford Motor Co.*, 848 F.3d 219, 228 (4th Cir. 2017) (internal quotation marks and citations omitted).

We turn then to consider the district court’s decision to latch *Daubert*’s gate and exclude the testimony of three of the plaintiffs’ expert witnesses.⁴

⁴ We do not address at any length the qualifications of these experts, for it is clear from the record that plaintiffs’ experts are all quite qualified, and Pfizer does not assail their credentials. We therefore focus our discussion on the substance of their testimony.

A. Dr. Jewell (Statistics)

The plaintiffs offered the testimony of Dr. Nicholas Jewell to establish a statistical association between Lipitor and new-onset diabetes. To reach his conclusion, Dr. Jewell performed a reanalysis of the data used in several clinical trials involving Lipitor, including the Pfizer-sponsored ASCOT-LLA trial,⁵ as well as clinical trial data submitted to the FDA in the NDA for Lipitor. The district court excluded Dr. Jewell's opinions as to the ASCOT and NDA data, and the plaintiffs challenge both exclusions on appeal.

1. NDA Data

Before a drug may be sold or marketed in the United States, its manufacturer must submit an NDA to the FDA for review and approval. *See* 21 CFR § 314.1. A typical NDA includes a clinical data selection, which contains an overview of clinical investigations of the drug. *See id.* § 314.50(d)(5). In his initial expert report, Dr. Jewell reviewed data underlying seven clinical trials submitted as part of the NDA for Lipitor. Dr. Jewell examined blood glucose levels of patients in the trials who had taken Lipitor as compared to those who had taken a placebo. Persistently elevated blood glucose levels may, but do not necessarily, indicate that a patient is diabetic.

At the outset of his report, Dr. Jewell confessed that the data provided “less than optimum information about [Lipitor’s] effect on glucose metabolism or new-onset

⁵ *See* Peter S. Sever et al., *Prevention of Coronary and Stroke Events with Atorvastatin in Hypertensive Patients Who Have Average or Lower-than-Average Cholesterol Concentrations, in the Anglo-Scandinavian Cardiac Outcomes Trial—Lipid Lowering Arm (ASCOT-LLA): A Multicentre Randomised Controlled Trial*, 361 *Lancet* 1149 (2003).

diabetes, because of their short duration, relatively small sample sizes, and the unusual imbalance between the number of participants allocated to placebo and atorvastatin treatment.” CMO 54, 145 F. Supp. 3d 573, 577–78 (D.S.C. 2015) (quoting Dr. Jewell’s report). Nevertheless, Dr. Jewell proceeded to analyze the data and concluded that it reasonably showed that Lipitor had a statistically significant effect on baseline glucose metabolism. That association, he opined, should have put Pfizer on notice of a possible association between atorvastatin and diabetes.

The district court found Dr. Jewell’s methodology and application too tainted with potential bias and error to pass *Daubert* muster. First, the court expressed concern about the patients whose data were included in the sample set. In particular, Dr. Jewell included patients who had only a single elevated blood glucose reading and patients who had an abnormally elevated baseline glucose level at the beginning of the trial. Including patients with only a single reading, the district court explained, was contrary to the position the plaintiffs and their experts had taken elsewhere in the litigation (including through testimony from Dr. Jewell himself) that a single elevated glucose reading was not a reliable indicator of a diabetic patient. And including patients with elevated baseline levels risked introducing confounding variables into the analysis and thereby endangered the reliability of the results.⁶

⁶ As the district court explained, confounding variables “are those that correlate with the independent and dependent variable” and “cause a correlation to exist” between those variables “without causation being present.” CMO 54, 145 F. Supp. 3d at 580 n.11 (citing Federal Judicial Center, Reference Manual on Scientific Evidence 219, 285 (3d ed. 2011)). Here, “[t]here may be more participants with elevated on-treatment glucose levels in the

The plaintiffs argue that Dr. Jewell was merely opining as to an increased risk of elevated glucose levels rather than a specific increase in the risk of diabetes. But that argument is at odds with Dr. Jewell's own report, which concludes that the NDA "should have alerted . . . Pfizer to the possibility of increased risk of *new-onset diabetes* associated with atorvastatin treatment." J.A. 2254 (emphasis added).

Moreover, the district court believed that Dr. Jewell, a statistician, was simply not qualified to make determinations about which patients' data should have alerted Pfizer to a possible association between its drug and diabetes. Given that Dr. Jewell "readily admitted that he had no expertise in diabetes, did not 'quite know' what new-onset diabetes meant, and was unwilling to testify about the role or use of blood glucose as a surrogate marker for diabetes because he was not a clinician," the district court found that he "by his own testimony . . . lacked the expertise to opine about any implications that single glucose readings might have about the possibility of new-onset diabetes." CMO 67, MDL No. 2:14-mn-02502-RMG, 2016 WL 827067, at *2 (D.S.C. Feb. 29, 2016).

Also casting a specter of unreliability over Dr. Jewell's report was the manner in which he applied otherwise reliable statistical tests to the data. Statisticians rely on a range of mathematical tests to extrapolate meaning from data. Choosing the test to apply is a matter of selecting the appropriate tool for the task, and involves consideration of a variety of factors including type of data and sample size.

Lipitor group, not because the Lipitor caused elevated glucose levels, but because there were more participants with elevated glucose to start with in the Lipitor group." *Id.*

The district court questioned the way in which Dr. Jewell chose to include (and omit) the results from different tests in his report. Specifically, the district court noted that Dr. Jewell originally performed a “Fisher’s exact test” on the data, which is a “[w]ell-known small-sample technique” used to compare two samples. *See* CMO 54, 145 F. Supp. 3d at 583; Federal Judicial Center, Reference Manual on Scientific Evidence 255 n.108 (3d. ed. 2011) (“RMSE”). The Fisher’s exact test compares all possible outcomes in a distribution against the results actually observed to determine whether there is a statistically significant association. Generally speaking, an association is considered statistically significant if its p-value is less than 0.05. The p-value represents the likelihood that an apparent association observed in a data set is the product of random chance rather than a true relationship. Thus, when a p-value is less than 0.05, there is a less than 5% chance that a discovered association is a false positive result.⁷

When Dr. Jewell first performed the test, it showed a statistically insignificant association between Lipitor and the onset of diabetes, yielding a p-value of .0654. *See* CMO 54, 145 F. Supp. 3d at 582–83. Dr. Jewell, however, omitted that result from his initial expert report, and instead reported the results using a mid-p value,⁸ which returned a statistically significant result of $p=.04$. *See id.* at 583.

⁷ We discuss p-values and statistical significance at length in Section II.B.2., *infra*.

⁸ The mid-p value is “based on the same underlying mechanics and probability calculations as the [Fisher’s exact test], but it includes a correction to reduce the conservativeness and to provide a better balance of Type I and Type II errors.” Dan A. Biddle & Scott B. Morris, *Using Lancaster’s Mid-P Correction to the Fisher’s Exact Test for Adverse Impact Analyses*, 96 J. Applied Psychol. 956, 958 (2011). A Type I error

The district court explained that “using the mid-p approach, standing alone, does not render Dr. Jewell’s analysis unreliable,” and acknowledged that the test “is used by some statisticians and can be a valid methodology.” *Id.* Instead, the court was concerned that Dr. Jewell’s test selection was “results driven,” and that he “only used [the mid-p] test once the Fisher exact test returned a non-significant result.” *Id.*

Result-driven analysis, or cherry-picking, undermines principles of the scientific method and is a quintessential example of applying methodologies (valid or otherwise) in an unreliable fashion. “[C]ourts have consistently excluded expert testimony that ‘cherry-picks’ relevant data,” *EEOC v. Freeman*, 778 F.3d 463, 469 (4th Cir. 2015) (collecting cases), because such an approach “does not reflect scientific knowledge, is not derived by the scientific method, and is not ‘good science.’” *In re Bextra & Celebrex Mktg. Sales Practices & Prods. Liab. Litig.*, 524 F. Supp. 2d 1166, 1176 (N.D. Cal. 2007).

The plaintiffs defend the statistical legitimacy of the mid-p test. They also say that Dr. Jewell didn’t cherry-pick his methods but rather performed a variety of calculations on the data. But the district court didn’t take issue with the use of the mid-p test itself. Rather, it objected to Dr. Jewell’s choice to include in his report the results of some tests he performed (which supported the plaintiffs’ argument) but exclude the results of another

represents a “false positive” or “false alarm” error in which the null hypothesis (here, that there is no association between Lipitor and diabetes) is falsely rejected. *See* RMSE at 251 n.100. A Type II error is a “false negative,” or a false acceptance of the null hypothesis. *Id.* at 254 n.106.

(which did not). The district court concluded (we think reasonably) that such an approach lacked the hallmark of science properly performed.

Finally, the district court was skeptical of Dr. Jewell's use of average blood glucose increase as a metric in his analysis. Dr. Jewell calculated the average increase in blood glucose among patients with elevated levels in either arm of the trials (i.e., patients who had taken Lipitor or patients who had taken a placebo). He then reasoned that this average increase could be applied to the group of patients who had taken Lipitor, using his past (potentially flawed) conclusions to support the proposition that patients with glucose elevation were more likely to be in the Lipitor (rather than placebo) groups.

The district court found this logic to be "convoluted, and flawed." CMO 54, 145 F. Supp. 3d at 585. Because of the court's other issues with Dr. Jewell's process, it was understandably worried that the assumptions used in this part of the analysis were inherently unsound. Additionally, the court noted that the data actually showed greater increases in blood glucose among placebo-taking patients (who averaged a 37 mg/dL⁹ increase) compared to those on Lipitor (27.1 mg/dL). *Id.* "Thus," the court explained, "to the extent one can infer anything from the average glucose increases in this incredibly limited data set of 40, those who experienced meaningful changes in their glucose level had lower increases if taking Lipitor." *Id.* (internal quotation marks and footnote omitted). Consequently, the district court found that this data was unreliable, and in any event had

⁹ Measurements of blood glucose can be expressed in terms of the milligrams of glucose per deciliter of blood (mg/dL).

real potential to confuse jurors. Thus, even if Rule 702 did not bar admission of the opinion, the court would exclude it under Rule 403 because it was more likely to confuse the jury than to impart probative information. *See* CMO 67, 2016 WL 827067, at *5.

These are classic concerns regarding reliability and relevance that district courts should weigh when deciding whether to admit expert evidence. The court here properly discharged its gatekeeping duty by considering—and ultimately excluding—Dr. Jewell’s opinions, and explaining in detail its well-reasoned grounds for doing so.

2. *ASCOT-LLA*

Dr. Jewell also performed a reanalysis of data Pfizer produced from ASCOT-LLA, a clinical trial of Lipitor.¹⁰ In the trial, patients with certain cardiovascular risk factors were randomly assigned to one of two arms: one group of patients was prescribed a dose of Lipitor; the remaining patients received a placebo. The primary endpoints of ASCOT-LLA (that is, the outcomes for which study investigators were principally monitoring) were nonfatal myocardial infarction (heart attack) and fatal coronary heart disease. The trial was ultimately terminated early because Lipitor was proving so effective at reducing cardiovascular events that the data safety monitoring board recommended—and the steering committee agreed—to conclude the study ahead of its originally scheduled timeline.

¹⁰ Notably, Dr. Jewell didn’t include an analysis of the ASCOT study in his original expert report, but the district court permitted him to file a supplemental report to address it.

Development of diabetes was a tertiary endpoint in ASCOT, which (like all identified endpoints) was monitored through ASCOT's adjudication process. Study investigators sent patient information relevant to any endpoint to a special endpoints committee. That committee applied previously-defined criteria for classifying diagnoses to determine whether a particular endpoint had been reached.¹¹ The data the committee received was blinded, meaning that its members weren't told whether an individual patient had been assigned to the atorvastatin or placebo arm of the trial. After it determined whether a patient had reached a designated endpoint, the committee reported its findings to the coordinating center, which in turn communicated them to the study's data safety monitoring board.

The ASCOT-LLA authors ultimately reported no statistically significant difference between the rate at which patients who took Lipitor developed diabetes and the rate among the patients who received a placebo. "In other words," the district court summarized, "the study did not find an association between Lipitor and new-onset diabetes." CMO 54, 145 F. Supp. 3d at 589. Dr. Jewell, however, chose to reanalyze the underlying data and reached an entirely opposite conclusion: he found that the data showed that Lipitor use was

¹¹ The ASCOT Endpoint Manual used by the endpoints committee set forth criteria for diagnosing diabetes based on guidance from the World Health Organization. Those criteria included patients who exhibited either "(i) Fasting plasma glucose ≥ 7.0 mmol/l on two occasions," "(ii) 2 hour post 75g glucose load plasma glucose ≥ 11.1 mmol/l," or "(iii) Unequivocal hyperglycemia with acute metabolic decompensation or obvious symptoms." CMO 54, 145 F. Supp. 3d at 590 (internal quotation marks omitted). The unit mmol/l indicates millimoles per liter and is another way to express a blood glucose measurement.

in fact associated with a significantly increased risk of new-onset diabetes among patients at risk for the disease.

How did Dr. Jewell reach this contrary result? The primary juncture where the ASCOT authors and Dr. Jewell part ways involves selection of the proper criteria used to diagnose a patient with diabetes. According to Dr. Jewell, it was unclear from the ASCOT protocol how investigators were identifying diabetic patients. He also quibbled with the criteria for diagnosing patients with diabetes set forth in the ASCOT protocol. Additionally, Dr. Jewell determined that reanalysis was appropriate because he could not reconstruct the study's adjudication process based on the data Pfizer had produced. Analyzing the data anew using his preferred definition of new-onset diabetes¹² (and without the advantage of independently adjudicated data), Dr. Jewell concluded that the ASCOT-LLA data indeed showed an association between use of Lipitor and an increased risk of diabetes.

The district court again took issue with Dr. Jewell's methodology. First, the court expressed concern with Dr. Jewell's decision to replace the definition of diabetes used by the ASCOT endpoints committee with one of his own. Although Dr. Jewell is well-qualified as a statistician, he's not a medical doctor or professional, nor does he have any particular expertise in diabetes. The court decided that Dr. Jewell lacked the expertise to

¹² Dr. Jewell included only patients who exhibited "two or more on-treatment glucose values > 125 mg/dL," a "subset of the definition used by the Endpoints Committee" and the "first of the three criteria in the [World Health Organization] definition." CMO 54, 145 F. Supp. 3d at 591.

“second guess” the judgments of the endpoints committee, and that it was inappropriate for “someone with no clinical expertise [to choose] to replace the adjudication committee’s determination of new-onset diabetes with particular unadjudicated raw data, namely lab values of his choice.” CMO 54, 145 F. Supp. 3d at 591. The district court stressed the importance of using prespecified diagnostic criteria for identifying endpoints (as the ASCOT endpoints committee had) to “help[] guard against bias.” *Id.* at 590. By contrast, Dr. Jewell, who confessed confusion over the precise criteria the ASCOT investigators had used to diagnose diabetes, had the benefit of hindsight when he selected his definition.

Moreover, despite using an apparently narrower definition of diabetes than the one used by the endpoints committee, Dr. Jewell nevertheless discovered *more* cases of new-onset diabetes. The court expressed skepticism about the way in which Dr. Jewell identified the cases because his decisions—unlike those of the adjudication committee in the ASCOT study—were not blind and thus lacked the same insulation from possible bias. And the endpoints committee (unlike Dr. Jewell) “had access to medical records and case files and examined and reconciled data to ensure that the inclusion of a case was accurate.” *Id.* (internal quotation marks and alterations omitted).

The plaintiffs say that Dr. Jewell was right to perform a reanalysis given certain “anomalies” that he identified in the original study. To be sure, a reanalysis of a clinical trial’s data may sometimes be appropriate. But such a reanalysis isn’t per se admissible under *Daubert*. As with any expert undertaking, a district court must review a reanalysis to ensure it meets *Daubert*’s demands of reliable methodologies reliably applied. And when the results of a reanalysis are squarely at odds with the conclusions of a published,

peer-reviewed study; the methods of the reanalysis are questionable because of the absence of properly adjudicated data; the expert performing the reanalysis lacks expertise in the portion of the study he's modifying; and that expert offers an unpersuasive rationale for why the original findings are wrong and his correct, then skepticism by the district court is warranted.

Here, the district court understandably worried that Dr. Jewell, whose expertise lay in the realm of mathematics rather than medicine, compromised the soundness of his opinions by eschewing the benefits of ASCOT's blinded adjudication process and usurping its predetermined diagnostic criteria with his own without adequate justification. Such a determination is well within the broad discretion we afford district courts in deciding evidentiary questions.

The plaintiffs make much of a few statements by the district court suggesting that Dr. Jewell substituted his definition of diabetes for that of the ASCOT investigators "without any explanation." *E.g.*, CMO 54, 145 F. Supp. 3d at 593. In fact, Dr. Jewell did express his belief that the definition used in the ASCOT protocol was flawed and risked underestimating the incidence of diabetes. But that's not sufficient to call into question the district court's understanding of Dr. Jewell's approach. Read in context, the court repeatedly opined that Dr. Jewell (1) wasn't qualified to second-guess the clinicians who selected the ASCOT diagnostic criteria; (2) didn't even know what criteria they had used; and (3) hadn't adequately justified his decision to replace their criteria with his own.

"The touchstones for admissibility under *Daubert* are two: reliability and relevancy." *United States v. Crisp*, 324 F.3d 261, 268 (4th Cir. 2003). Here, the district

court determined that Dr. Jewell’s scientific approach was flawed and thus unreliable, and it clearly and cogently explained why. We emphasize again the inherently “flexible” nature of the *Daubert* analysis. *Daubert*, 509 U.S. at 594. “Many factors will bear on the inquiry,” and there is no “definitive checklist or test.” *Id.* at 593. In some cases, no individual methodological or application discrepancy may be sufficient to independently justify exclusion of an expert’s testimony. Nevertheless, courts must look to the entire process that produced an opinion to determine whether the expert’s work satisfies *Daubert*’s fundamental command: that expert testimony be reliable and relevant.

That is what the district court did here—it identified and articulated clear (and, we think, reasonable) concerns it had about the manner in which Dr. Jewell reached his conclusions. Even when afforded the opportunity to supplement his report, Dr. Jewell failed to assuage the court’s justified worries with his testimony. Accordingly, we affirm the district court’s ruling excluding Dr. Jewell’s expert opinions.

B. Dr. Singh (General Causation)

The plaintiffs next ask us to revive the testimony of Dr. Sonal Singh, who offered general causation opinions. Dr. Singh conducted a literature review and performed a meta-analysis of relevant studies. Ultimately, he concluded that there was an association between the use of Lipitor and increased risk of diabetes. Dr. Singh then applied what epidemiologists refer to as the “Bradford Hill criteria” to the data to determine whether the association was a causal one; that is, whether use of Lipitor in fact caused patients’ increased risk of diabetes.

The Bradford Hill criteria are a series of factors used by epidemiologists to determine whether an observed association between two variables is causal. Those factors include temporal relationship, strength of the association, dose-response relationship, replication of the findings, biological plausibility, consideration of alternative explanations, cessation of exposure, specificity of the association, and consistency with other knowledge. *See* RMSE at 599–600. No single criterion is dispositive: a true causal relationship may exist where one or more factors are absent; likewise, one or more factors may be present in a dataset where there is no causal relationship at all. *Id.*

Dr. Singh considered the Bradford Hill criteria and concluded that there was indeed a causal relationship between Lipitor and diabetes. But after Dr. Singh submitted his initial report, the district court ruled that he (and other experts) “must demonstrate . . . that particular doses of Lipitor are capable of causing diabetes,” and directed him to file a supplemental report “offering opinions as to whether Lipitor causes diabetes at” each of the four doses commercially available in the United States (10, 20, 40, and 80 mg). *See* CMO 49, MDL No. 2:14-mn-02502-RMG, 2015 WL 6941132, at *6 (D.S.C. Oct. 22, 2015). Dr. Singh prepared a supplemental report, in which he concluded that the data did indeed show a causal association between Lipitor and diabetes at each commercial dose. Nevertheless, the district court excluded Dr. Singh’s opinion for each dose except 80 mg.

The district court thought it improper for Dr. Singh to apply the Bradford Hill criteria to the data for the 10 mg dose of Lipitor to determine causality because the data did not reveal a sufficiently strong association between the drug and the disease. By Dr. Singh’s own admission, his opinions as to the 20 and 40 mg doses were based on his

conclusions for the 10 mg dose, and thus without Dr. Singh's opinion as to the 10 mg dose, the district court held that his opinions as to the 20 and 40 mg doses must also be excluded. But the data did show a sufficient relationship between an 80 mg dose of Lipitor and diabetes, and because Dr. Singh's application of the Bradford Hill criteria suggested that the relationship there was a causal one, the district court denied Pfizer's motion to exclude Dr. Singh's opinions with respect to the 80 mg dose.

The plaintiffs raise two challenges to these rulings: they argue that the district court erred in requiring Dr. Singh to (1) offer an opinion on causation at each specific dose of Lipitor; and (2) find a statistically significant association between Lipitor and diabetes before opining as to causation. Neither argument is persuasive.

1. Dosage

“[D]ose matters.” So explained the court in *In re Bextra & Celebrex Mktg. Sales Practices & Prod. Liab. Litig.*, 524 F. Supp. 2d 1166, 1174 (N.D. Cal. 2007), and so held the district court here. One of the central tenets of toxicology is that “the dose makes the poison.” RMSE at 636. “[T]his implies that all chemical agents are intrinsically hazardous,” and that “whether they cause harm is only a question of dose.” *Id.*

The district court determined that this case called for expert opinions broken down by dose, particularly because “where [as here] the experts agree that there is a dose-response relationship and where there is evidence that an association no longer holds at low doses, dose certainly matters, and Plaintiffs must have expert testimony that Lipitor causes, or is capable of causing, diabetes at particular dosages.” CMO 49, 2015 WL 6941132, at *6. This makes sense: in many cases, as the Reference Manual explains,

substances that might be quite harmful in high doses are innocuous in smaller amounts. *See* RMSE at 636. Indeed, “[e]ven water, if consumed in large quantities, can be toxic.” *Id.*

In requiring Dr. Singh to provide dose-specific opinions, the district court relied on *Westberry v. Gislaved Gummi AB*, 178 F.3d 257 (4th Cir. 1999). In that case, Westberry alleged he was injured when he used the defendant’s rubber gaskets, which had been coated with a layer of talcum powder before shipping. *Id.* at 260. Westberry claimed that he inhaled some of the talc coating, which resulted in injury to his sinuses. *See id.*

We explained that “[i]n order to carry the burden of proving a plaintiff’s injury was caused by exposure to a specified substance, the plaintiff must demonstrate the levels of exposure that are hazardous to human beings generally as well as the plaintiff’s actual level of exposure.” *Id.* at 263 (internal quotation marks omitted). Nevertheless, we recognized that “only rarely are humans exposed to chemicals in a manner that permits a quantitative determination of adverse outcomes,” and that it is often “difficult, if not impossible, to quantify the amount of exposure.” *Id.* at 264 (alteration and internal quotation marks omitted); *see also* CMO 49, 2015 WL 6941132, at *5 (acknowledging that “it is often difficult to establish with quantitative precision the specific level of exposure”). Thus in *Westberry*, it was sufficient for the plaintiff to introduce evidence of “substantial exposure.” *See* 178 F.3d at 264.

Pharmaceuticals like Lipitor, on the other hand, lend themselves quite well to dosage analysis. Unlike substances in other toxic tort cases (talcum powder, for instance, or asbestos), pharmaceutical drugs are typically prescribed and consumed in measured and

knowable quantities. Because patients on Lipitor know their precise dose, and because there is data available on many other patients taking that same dose, pharmaceutical injury litigation may indeed present the “rare” case *Westberry* described in which a patient is exposed to a chemical in a readily quantified way.

We do not suggest that every case involving a claim of injury resulting from a pharmaceutical drug will require a dose-by-dose analysis, and an expert witness will not necessarily need to define the precise lower bound of exposure risk. The appropriate level of analysis will depend on the circumstances of the case and the capacity of current scientific methods. But where, as here, each plaintiff took one of only several commercially available doses, clinical data exist that enable an expert to perform a causation analysis at each dose, and experts (including plaintiffs’ own) acknowledge that there is some relationship between dosage and harm, the district court doesn’t abuse its discretion in asking the expert to produce a dose-by-dose analysis.

2. *Statistical Significance*

Dr. Singh’s supplemental expert report accounted for the four commercial doses of Lipitor and again concluded that there was a causal association between the drug and diabetes at each particular dose. As before, Dr. Singh looked at clinical data to determine whether an association existed between the variables and then applied the Bradford Hill criteria to conclude that the observed relationship was causal.

Nevertheless, the district court excluded his opinion as to the association between diabetes and the 10 mg Lipitor dose. The court found it inappropriate for Dr. Singh to apply the Bradford Hill criteria where the association observed at the first stage of his

analysis was insufficiently strong. The Reference Manual on Scientific Evidence stresses that it is proper to employ the Bradford Hill criteria “only *after* a study finds an association to determine whether that association reflects a true causal relationship.” RMSE at 598–99. In fact, the Manual highlights cases in which “experts attempted to use these guidelines to support the existence of causation in the absence of any epidemiologic studies finding an association,” but observes that while “[t]here may be some logic to that effort . . . it does not reflect accepted epidemiologic methodology.” *See id.* at 599 n.141; *see also In re Fosamax Prods. Liab. Litig.*, 645 F. Supp. 2d 164, 188 (S.D.N.Y. 2009) (“Several courts that have considered the question have held that it is not proper methodology for an epidemiologist to apply the Bradford Hill factors without data from controlled studies showing an association.”); *Dunn v. Sandoz Pharm. Corp.*, 275 F. Supp. 2d 672, 678 (M.D.N.C. 2003) (rejecting expert contention that “it is not necessary to have an epidemiological study that demonstrates an association as a prerequisite for applying the Bradford Hill criteria”).

This case is different in that Dr. Singh did claim to discover an association before applying the Bradford Hill criteria. But the district court found that Dr. Singh couldn’t reliably apply the Bradford Hill criteria to the data because the association he found was too weak. How strong is strong enough? “[I]t is well established,” the district court explained, “that the Bradford Hill method used by epidemiologists . . . require[s] that an association be established through studies with statistically significant results.” CMO 68, 174 F. Supp. 3d 911, 924–25 (D.S.C. 2016) (collecting cases). Because Dr. Singh failed to find a statistically significant relationship between Lipitor and diabetes (and thereby

could not satisfy the first step of the Bradford Hill analysis), he should not have applied the factors at the second step. Doing so, the court reasoned, resulted in an unreliable opinion that warranted exclusion.

We pause here to provide a brief overview of the concept of statistical significance and its proper role in the courtroom. Statistical significance is a measure of confidence that a trend observed in a dataset is not random. “A study that is statistically significant has results that are unlikely to be the result of random error” RMSE at 573. Statistical significance is typically expressed through a p-value. “A p-value represents the probability that an observed positive association could result from random error even if no association were in fact present.” *Id.* at 576 (emphasis removed). To determine whether an association is statistically significant, statisticians compare the p-value to a predetermined threshold value (also known as a significance level). If the p-value is smaller than the significance level, then the finding is statistically significant. Otherwise, it is not. “The most common significance level . . . used in science is .05. A .05 value means that the probability is 5% of observing an association at least as large as that found in the study when in truth there is no association.” *Id.* at 577 (footnote omitted).

The purpose of statistical significance, then, is to indicate a certain level of confidence in the results of an analysis. A significant p-value is not, however, some all-purpose salve, nor is it a get-out-of-*Daubert*-free card. Just as statistically significant evidence won’t result in automatic admission, the absence of a p-value that is smaller than .05 (or some other threshold) isn’t necessarily fatal to a case. Rather, statistical significance may bear on the question of reliability, and must therefore be subjected to the same inquiry

as any other scientific evidence—including whether the expert has applied the “same level of intellectual rigor that characterizes the practice of an expert in the relevant field.” *Kumho Tire*, 526 U.S. at 152.

The Supreme Court addressed this issue in *Matrixx Initiatives, Inc. v. Siracusano*, 563 U.S. 27 (2011). There, it considered a claim that a company had committed securities fraud by failing to disclose an alleged risk of one of its pharmaceutical products, despite the fact that the information the company had received didn’t show “a statistically significant number of adverse events.” *Id.* at 30. The question there was not about expert testimony admissibility; instead, the Court faced the issue of whether statistically insignificant data could be material, thus potentially requiring disclosure to shareholders. *See id.* The Court declined to “adopt a bright-line rule that reports of adverse events associated with a pharmaceutical company’s products cannot be material absent a sufficient number of such reports to establish a statistically significant risk that the product is in fact causing the events.” *Id.* at 39 (footnote omitted).

“A lack of statistically significant data,” it explained, “does not mean that medical experts have no reliable basis for inferring a causal link between a drug and adverse events.” *Id.* at 40. Indeed, the Court observed that “medical experts rely on other evidence to establish an inference of causation” and that “courts frequently permit expert testimony on causation based on evidence other than statistical significance.” *Id.* at 40–41 (expressly declining to answer “whether the expert testimony was properly admitted in those cases” or “define . . . what constitutes reliable evidence of causation”). Finally, the Court noted that neither medical professionals nor the FDA rely solely on statistically significant

evidence when evaluating a drug’s safety, and instead consider “a wide range of evidence of causation.” *Id.* at 41–42.

We think *Matrixx* is consistent with *Daubert*, which teaches that judges should look beyond the confines of the courtroom to ask what experts do in the real world. And the record, literature, and case law make clear that statistical significance, while important in many contexts, doesn’t always paint a full portrait. Thus, we decline to establish a bright-line rule requiring experts to rely only on evidence that is statistically significant or else have their opinions excluded.

The district court here did not hold to the contrary. Instead, the court focused on the particular analysis Dr. Singh performed—applying the Bradford Hill criteria to a set of data—and determined that in that specific context, the analysis requires a statistician to find a statistically significant association at step one before moving on to apply the factors at step two. Because Dr. Singh deviated from this norm, the district court concluded that he unreliably applied the Bradford Hill methodology and thus his opinion was subject to exclusion.

Had the district court held (as the plaintiffs fret) that statistically significant evidence is a sine qua non of admissibility, it might well have committed an abuse of discretion. But the court here did no such thing. Instead, it excluded Dr. Singh’s opinion as to the 10 mg dose because the plaintiffs “failed to demonstrate that Dr. Singh’s reliance on non-statistically significant ‘trends’ is accepted in his field, that non-statistically significant findings have served as the basis for any epidemiologist’s causation opinion in peer-reviewed literature, or that standards exist for controlling the technique’s operation.”

CMO 68, 174 F. Supp. 3d at 926. “These *Daubert* factors,” the district court explained, “all suggest a lack of reliability” in Dr. Singh’s opinions. *Id.* Such a conclusion was well within the district court’s discretion.

By Dr. Singh’s own admission, without his opinion that there is an increased risk of diabetes at the 10 mg dose of Lipitor, his opinions as to the 20 and 40 mg doses must also fall. As the district court explained, Dr. Singh’s conclusions regarding the 20 and 40 mg doses were “an inference” based on his findings regarding the 10 and 80 mg doses. *Id.* at 927 (quoting Dr. Singh as finding it “difficult to imagine how atorvastatin 10 mg and 80 mg can increase the risk of diabetes without similar risk seen with atorvastatin 20 mg and 40 mg”).

We thus find no abuse of discretion in the district court’s decision to exclude Dr. Singh’s general causation opinions at the 10, 20, and 40 mg doses of Lipitor.

C. Dr. Murphy (Specific Causation)

The final expert witness at issue in this appeal is Dr. Elizabeth Murphy. The plaintiffs sought to use Dr. Murphy to prove specific causation in the bellwether case of Juanita Hempstead.¹³ “For specific causation, the plaintiff must ‘demonstrate that the

¹³ Ms. Hempstead used a 20 mg dose of Lipitor. Given our affirmance of the district court’s other rulings in this case (including the exclusion of the plaintiffs’ general causation testimony for doses less than 80 mg), summary judgment in Ms. Hempstead’s case would be proper even if Dr. Murphy’s analysis could survive Pfizer’s challenge. *See Westberry*, 178 F.3d at 263–64 (proof of general and specific causation required). Proving this case requires evidence of general causation, and the plaintiffs have not argued (nor, we think, could they) that Dr. Murphy’s testimony satisfies that burden. Without such evidence, Ms. Hempstead’s case cannot stand. Nevertheless, because it provides an independent basis

substance actually caused injury in her particular case.’” CMO 55, 150 F. Supp. 3d 644, 649 (D.S.C. 2015) (quoting *Guinn v. AstraZeneca Pharm. LP*, 602 F.3d 1245, 1249 n.1 (11th Cir. 2010)) (alteration omitted). In other words, the focus of Dr. Murphy’s testimony wasn’t whether Lipitor can cause diabetes generally, but whether Lipitor actually caused Ms. Hempstead to develop the disease.

To reach her result, Dr. Murphy considered (1) the existence of “reports in the scientific literature and/or reliable expert data analyses showing the occurrence of new onset diabetes with Lipitor”; (2) whether it was “biologically plausible that Lipitor can cause new onset diabetes”; (3) whether “new onset diabetes appear[ed] after the Lipitor was given”; (4) the existence of “other possible causes for the development of new onset diabetes”; and (5) the likelihood that “Lipitor caused new onset diabetes in this individual at this time.” CMO 55, 150 F. Supp. 3d at 652. Based on her analysis, Dr. Murphy concluded that Ms. Hempstead’s use of Lipitor was a substantial contributing factor in her development of diabetes.

The district court expressed skepticism regarding Dr. Murphy’s analysis, noting that while her report claimed that the methodology is “commonly used,” Dr. Murphy “could not identify any organizations or peer-reviewed texts that contain this methodology,” nor any colleagues who used the methodology to determine the cause of diabetes. *Id.* Additionally, Dr. Murphy had not personally used the methodology to determine the cause

for the disposition of Ms. Hempstead’s claim, we explain why the district court’s exclusion of Dr. Murphy’s testimony on specific causation was not an abuse of discretion.

of her own patients' diabetes and had never diagnosed one of her patients as suffering from statin-induced diabetes. *Id.*

The district court also found that Dr. Murphy's causation analysis rested primarily on the first three steps, which relate to general causation and a temporal relationship. *See id.* at 652–53. Although Dr. Murphy did identify and discuss other risk factors at step four of her methodology, the district court held that she had to do more than merely identify and summarily dismiss alternative possible causes. *See id.* at 659–61. “Under Fourth Circuit law,” the court explained, “an expert need not rule out every possible alternative cause of a disease in a differential diagnosis. However, she must offer an explanation as to why these other recognized causes, alone, are not responsible for the disease in a particular plaintiff.” *Id.* at 661 (citation omitted). The court thus rejected Dr. Murphy's analysis, finding that she “offers no data or facts to make the leap from a possibility to a probability that Lipitor was a substantial contributing factor.” *Id.*

The plaintiffs urge that the district court erred in excluding Dr. Murphy's testimony because differential diagnoses are a reliable, widely used technique to determine whether a particular source is the likely cause of an individual's disease. Indeed, we have explained that a differential diagnosis is “a standard scientific technique of identifying the cause of a medical problem by eliminating the likely causes until the most probable one is isolated.” *Westberry*, 178 F.3d at 262. Such an analysis is accomplished by “determining the possible causes for the patient's symptoms and then eliminating each of these potential causes until reaching one that cannot be ruled out or determining which of those that cannot be excluded is the most likely.” *Id.*

A rose by another name may smell as sweet—but simply calling an analysis a differential diagnosis doesn't make it so. See *McClain v. Metabolife Intern., Inc.*, 401 F.3d 1233, 1253 (11th Cir. 2005) (“[A]n expert does not establish the reliability of his techniques or the validity of his conclusions simply by claiming that he performed a differential diagnosis on a patient.”); *Black v. Food Lion, Inc.*, 171 F.3d 308, 313–14 (5th Cir. 1999). And “[n]ot every opinion that is reached via a differential-diagnosis method will meet the standard of reliability required by *Daubert*.” *Best v. Lowe’s Home Ctrs., Inc.*, 563 F.3d 171, 179 (6th Cir. 2009). “Although a reliable differential diagnosis need not rule out all possible alternative causes, it must at least consider other factors that could have been the sole cause of the plaintiff’s injury.” *Guinn v. AstraZeneca Pharm. LP*, 602 F.3d 1245, 1253 (11th Cir. 2010); see also *Westberry*, 178 F.3d at 265 (“A differential diagnosis that fails to take serious account of other potential causes may be so lacking that it cannot provide a reliable basis for an opinion on causation.”).

In this case, Dr. Murphy did consider (and purportedly ruled out) several other risk factors, including Ms. Hempstead’s family history, race, body mass index (“BMI”), and age. But her analysis of those factors—and, more importantly, her reasons for rejecting them as the likely cause of Ms. Hempstead’s disease—fell short. As the district court recognized, Dr. Murphy identified several factors for which the risk “greatly exceed[ed] the risk of developing diabetes associated with Lipitor.” CMO 55, 150 F. Supp. 3d at 661. Dr. Murphy opined in her report that several of Ms. Hempstead’s risk factors were significant to the development of her disease and others were not. Nevertheless, Dr.

Murphy concluded that Lipitor ingestion was a substantial contributing factor in causing Ms. Hempstead's diabetes.

The district court found "many difficulties" with Dr. Murphy's testimony, explaining that she failed to "employ a reliable methodology to determine that Lipitor was a substantial contributing factor in Plaintiff[']s development of diabetes." *Id.* The court explained: "The powerful evidence that Plaintiff's many other risk factors can independently cause diabetes and cannot be ruled out further undermine Dr. Murphy's testimony. The gap between the available scientific evidence and Dr. Murphy's opinions [is] too great to survive a Rule 702 review." *Id.*

The plaintiffs argue that Dr. Murphy did in fact explain why the other significant risk factors were less likely to have caused Ms. Hempstead's disease. They are correct that Dr. Murphy's report was not entirely void of explanation. For example, Dr. Murphy explained that the family history factor was likely attenuated given the large size of Ms. Hempstead's family and the limited number of family members who developed the disease. Dr. Murphy addressed the significant risk factors of age and BMI by explaining that studies either adjust for those factors or are randomized to control for them, and that those studies nevertheless show an increased risk of diabetes associated with Lipitor.

But these explanations (even if true) do not accomplish the specific causation expert's task: accounting for the development of the disease in a particular plaintiff. That Lipitor may cause an increased risk of diabetes notwithstanding certain other risk factors is insufficient to conclude that the drug was a substantial contributing factor in an individual patient. To hold otherwise would obviate the need for any specific causation

evidence at all. *Cf. Best*, 563 F.3d at 179 (when ruling out other causes in a differential diagnosis, “the doctor must provide a reasonable explanation as to why he or she has concluded that any alternative cause suggested by the defense was not the sole cause” (internal quotation marks and alteration omitted)). This fueled the district court’s unease that Dr. Murphy appeared to simply conclude that “so long as the patient took Lipitor and developed diabetes, then Lipitor was a substantial contributing factor.” CMO 55, 150 F. Supp. 3d at 661.

Dr. Murphy’s deposition testimony only bolstered this concern. Dr. Murphy testified that Ms. Hempstead’s BMI and adult weight gain were substantial contributing factors to her development of diabetes. *Id.* at 655. She also opined that if a “patient was taking the Lipitor and they developed diabetes while on it . . . I would think that it would be a contributing factor.” *Id.* at 652 (quoting deposition transcript). These statements buttressed the district court’s conclusion that Dr. Murphy hadn’t performed a reliable differential diagnosis because she focused primarily on the fact that the plaintiff took Lipitor close in time to her development of diabetes and also failed to adequately explain the basis for ruling out other contributing factors—including some that she herself described as substantial.

In sum, the court found that Dr. Murphy’s report appeared to dismiss other possible causes in favor of Lipitor in a cursory fashion that appeared closer to an ipse dixit than a reasoned scientific analysis. Dr. Murphy’s conclusions focused almost exclusively on the fact that Ms. Hempstead took the drug and later developed the disease, rather than

explaining what led her to believe that it was a substantial contributing factor as compared to other possible causes. Simply put, *Daubert* requires more.

As with the testimony of Drs. Jewell and Singh, the district court carefully subjected Dr. Murphy's testimony to *Daubert*'s twin rigors of relevance and reliability. The court acted well within its discretion in excluding Dr. Murphy's testimony.

III.

Left without sufficient expert testimony (at least for doses of Lipitor lower than 80 mg), the plaintiffs next appeal the district court's ruling that other evidence of causation was not enough to survive summary judgment. Pfizer argued—and the district court found—that absent admissible evidence of causation offered by the plaintiffs, there was no genuine dispute of a material fact, and thus summary judgment was appropriate. A district court properly grants summary judgment when “the movant shows that there is no genuine dispute as to any material fact and the movant is entitled to judgment as a matter of law.” Fed. R. Civ. P. 56(a). “We review a grant of summary judgment de novo, viewing all facts and inferences in the light most favorable to the nonmoving party.” *Balbed v. Eden Park Guest House, LLC*, 881 F.3d 285, 288 (4th Cir. 2018).

Here, the district court rejected the plaintiffs' argument that causation could be proven through purported admissions by Pfizer that Lipitor can cause diabetes, including an email from a Pfizer Senior Vice President, statements on the American and Japanese Lipitor labels, information in the Lipitor NDA, and a statement on the official Lipitor

website. The district court held that under *Erie*,¹⁴ state substantive law governs the means by which each plaintiff must prove her specific tort case. “To the extent that state substantive law requires causation to be established by expert testimony, it is also a question of state substantive law whether party-opponent admissions can substitute for expert evidence of causation.” CMO 100, 227 F. Supp. 3d 452, 469 (D.S.C. 2017). Because state courts “have not had an opportunity to pass on the specific question,” the district court attempted to predict what the highest-level state courts would decide if confronted with the issue. *See id.*

The district court rejected the plaintiffs’ contention that substantive state law on the issue of expert testimony meaningfully differs from one state to another. Instead, the court concluded that although “the specific language used by courts var[ies] to some degree, all jurisdictions require expert testimony at least where the issues are medically complex and outside common knowledge and lay experience.” *Id.*

The court supported this finding with a meticulous survey of authority spanning the law of the various states, territories, and the District of Columbia. *See id.* at 469–77. While the district court acknowledged that there are indeed “instances where expert testimony is not required to prove causation,” it explained that those circumstances, such as cases “where a lay juror can infer causation from common knowledge and lay experience—are not present here.” *Id.* at 477. Expert testimony would not be required, for example, “for a lay jury to determine that a gunshot wound to the head of an otherwise healthy person who

¹⁴ *Erie R.R. Co. v. Tompkins*, 304 U.S. 64 (1938).

died shortly thereafter was the proximate cause of her death.” *Cowart v. Widener*, 697 S.E.2d 779, 784 (Ga. 2010).

In a diversity case, state substantive law governs the “standard of care, whether it has been violated, and whether such violation is the proximate cause of plaintiff’s injury”—in other words, the “substantive elements” of a tort action. *Fitzgerald v. Manning*, 679 F.2d 341, 346 (4th Cir. 1982). We agree with the district court that it is substantive state law that must guide our decision here. For purposes of our analysis, we also assume (without deciding) that the plaintiffs’ evidence would be admissible at trial under the applicable rules of evidence. The issue, then, is “not so much whether the alleged admissions are admissible against” Pfizer. *In re Mirena IUD Prods. Liab. Litig.*, 202 F. Supp. 3d 304, 314 (S.D.N.Y. 2016). The question, rather, is “whether as a matter of substantive products liability law admissions can substitute for expert evidence of causation, given the widely held principle that expert testimony is required in cases involving a complex or technical question outside the ken of the average lay juror.” *Id.*

Our sister circuits have had only limited opportunity to weigh in on this issue. In *In re Meridia Prods. Liab. Litig.*, 447 F.3d 861 (6th Cir. 2006), the Sixth Circuit considered whether certain statements about the drug Meridia contained on its FDA labelling were sufficient to carry the burden at summary judgment. There, the court declined to hold that a warning contained in a drug’s label could “never create a triable issue of fact with respect to causation,” and instead affirmed the district court’s conclusion that “neither epidemiological nor expert evidence [was] necessary to a finding of causation.” *Id.* at 866.

But in so holding, the Sixth Circuit emphasized that the district court had relied on the “specific wording” of the label rather than the mere fact that the warning existed, and further stressed the “strong wording of the label,” which stated in no uncertain terms that the drug substantially increased its users’ blood pressure. *See id.* (contrasting the “strong language of ‘substantially increases’ with milder warning language such as ‘is associated with’”). What *Meridia* teaches us, then, is that in some circumstances a compelling admission may eliminate the need for expert evidence, but that determination depends in part on the character of the purported admission.

The Second Circuit considered this argument more recently in an unpublished decision in *In re Mirena IUD Prods. Liab. Litig.*, 713 F. App’x 11 (2d Cir. 2017). The posture of that case (which involved product liability claims against the manufacturer of a medical device) was quite similar to ours: the district court had excluded the plaintiffs’ causation experts and so the plaintiffs sought to survive summary judgment by introducing party admissions. *See id.* at 15–16. The district court had held (and the Second Circuit agreed) that “all fifty states typically require expert testimony to prove causation where the causal relationship is outside the common knowledge of lay jurors,” and that the plaintiffs had failed to identify a clear exception to that rule. *See id.* (citing our district court’s survey of authority).

The court avoided the ultimate question of “whether party admissions could ever substitute for expert testimony,” concluding instead that the purported statements, even if admitted, could not carry the burden to survive summary judgment because they were “simply not enough to establish general causation.” *See id.* (“[N]o reasonable juror could

find general causation more likely than not based on the Plaintiffs’ admissible evidence.”). We think that the independent sufficiency of the evidence here presents a closer question than in *Mirena*—but nevertheless, we agree with the district court that the evidence here is not enough to overcome summary judgment.

There may be cases involving complex issues in which a party admission standing alone can suffice to avoid summary judgment. But we would expect those cases to be rare indeed. *See Mirena*, 202 F. Supp. 3d at 315 (noting that “if admissions could ever substitute for expert testimony in a complex case that requires expert testimony as to causation under state law, those admissions would have to be clear, unambiguous, and concrete rather than an invitation to the jury to speculate as to their meaning”).

The questions presented in this case are complex and manifold (if the ink spilled in litigation thus far were not evidence enough). Moreover, the evidence at issue isn’t especially strong. As the district court explained, most of the statements don’t directly support the proposition that Lipitor causes diabetes, but instead speak to association rather than causation, or focus on blood glucose rather than diabetes. *See CMO 100*, 227 F. Supp. 3d at 481–85; *see also In re Zoloft (Sertralinehydrochloride) Prods. Liab. Litig.*, 176 F. Supp. 3d 483, 497–98 (E.D. Pa. 2016) (internal documents showing that employees “raised questions about associations between” drug and birth defects “may be relevant to questions of Pfizer’s knowledge and actions . . . but do not raise a genuine issue of material fact as to causation”).

To hand to a jury the evidence here and ask it to reach a conclusion as to causation with any amount of certainty would be farcical and would likely result in a verdict steeped

in speculation. Accordingly, the district court did not err in finding that the evidence could not independently establish a genuine dispute of a material fact.

IV.

Finally, the plaintiffs argue that it was improper for the district court to enter summary judgment across all cases in the MDL after no plaintiff produced adequate evidence of specific causation in response to the court's show cause orders. Instead, they argue, the scores of consolidated cases ought to have been returned to their respective transferor courts for individual resolution of the issue of specific causation.

The MDL statute permits the consolidation of several “civil actions involving one or more common questions of fact” for “coordinated or consolidated pretrial proceedings.” 28 U.S.C. § 1407(a). For consolidation to occur, the Judicial Panel on Multidistrict Litigation must conclude that transfer will serve “the convenience of parties and witnesses and will promote the just and efficient conduct of such actions.” *Id.* Each action transferred into an MDL “shall be remanded by the panel at or before the conclusion of such pretrial proceedings to the district from which it was transferred unless it shall have been previously terminated.” *Id.*

It is well established that a transferee court may dispose of cases in an MDL through summary judgment—and indeed, they often do. *See In re Food Lion, Inc. Fair Labor Standards Act Effective Scheduling Litig.*, 73 F.3d 528, 532 (4th Cir. 1996); *see also* Federal Judicial Center, *Manual for Complex Litigation* § 22.36 (4th ed. 2004) (“An MDL transferee judge has authority to dispose of cases on the merits—for example, by ruling on

motions for summary judgment . . .”). And although an MDL court may recommend that cases in an MDL be returned to their transferor courts, “the power to remand a transferred case rests with the [MDL] Panel and not with the transferee district judge.” *In re Roberts*, 178 F.3d 181, 184 (3d Cir. 1999). Here, rather than recommending to the Panel that the many cases in this MDL be resolved individually in their transferor courts by judges largely unfamiliar with the litigation and its complexities, the district court elected instead to resolve Pfizer’s summary judgment motion, resulting ultimately in dismissal of the plaintiffs’ claims.

We see no inconsistency between the district court’s actions and the MDL statute. The MDL device is intended to allow federal courts to “conserv[e] judicial resources in situations where multiple cases involving common questions of fact [are] filed in different districts.” *Food Lion*, 73 F.3d at 531–32. Transferee courts have broad discretion to determine whether ruling on a particular pretrial motion (such as summary judgment) is appropriate, or whether it would be better to ask the Panel to transfer cases back to their originating districts for individual resolution. Facing a motion for summary judgment, a court must consider whether the issues presented are well suited for the transferee court to deal with in a collective fashion. Where a motion for summary judgment “pertains to one or few cases, or rests on application of the transferor court’s conflicts-of-law and substantive law rules, the transferor judge may be able to decide the motions most efficiently. If the summary judgment motions involve issues common to all the cases centralized before the MDL court, however, the transferee judge may be in the best position to rule.” *Manual for Complex Litigation* § 22.36 (footnote omitted). *Principles of*

efficiency and fairness should guide the district court in deciding whether to resolve the motion or recommend remand.

Here, it was the district court's prerogative to determine whether it could dispose of the cases before it on the merits. At the time the court granted summary judgment, the plaintiffs were already facing an uphill battle: they were left without a general causation expert as to the majority of Lipitor doses, and their specific causation expert for the bellwether trial had also been excluded. "[T]he writing," as the district court put it, was "on the wall." CMO 100, 227 F. Supp. 3d at 491. After a series of show cause orders, no plaintiff had come forward with alternative evidence of specific causation that would survive review. The court explained that, after dealing with this litigation at great length, it was "familiar with the science and issues present and can dispose of the issues far more quickly and efficiently than dozens of courts spread across the country." *Id.* (stressing that it would be "inefficient, costly, and contrary to the purposes of the [MDL] statute to suggest remand without ruling on summary judgment"). The court thus declined to "essentially disregard the entire course of the MDL proceedings" by recommending remand. *Id.* (internal quotation marks omitted).

We find no error in the district court's decision to keep these cases and grant summary judgment to Pfizer.

V.

These cases involve difficult questions of mathematics and science, wrapped in a complex form of mass litigation. The district court here had to make a number of decisions pursuant to both its “gatekeeper” role imposed by *Daubert*, as well as its supervisory role as the transferee court in a large multidistrict litigation. The court discharged those duties meticulously and thoughtfully throughout the litigation, including performing careful review of the many expert reports, affording experts the opportunity to amend and revise those reports to ensure their opinions were fully considered, and, after ultimately excluding much of the plaintiffs’ expert testimony, allowing them to come forward with any additional evidence that could salvage what remained of their cases before rendering a final decision.

The district court’s judgments are therefore

AFFIRMED.