UNITED STATES COURT OF APPEALS

FOR THE SECOND CIRCUIT

August Term 2025
Argued: November 17, 2025 Decided: July 13, 2026

Docket Nos. 24-916-cv(L), 24-1121(Con), 24-2360(Con); 24-2594-cv

---

TIFFANY RUTLEDGE, INDIVIDUALLY AND AS MOTHER, GENERAL GUARDIAN OF, ET AL., KRISTOPHER WHITE, BRIDGET MCCONNELL, ALEXANDER HOLLAND, CHRISTINE HOLLAND,

Plaintiffs-Appellants,

v.

WALGREEN CO., COSTCO WHOLESALE CORPORATION, CVS HEALTH CORPORATION, CVS PHARMACY, INC., SAFEWAY INC., WALMART INC., A DELAWARE CORPORATION, RITE AID CORPORATION, FAMILY DOLLAR, INC., TARGET CORPORATION, SAM'S WEST, INC., DOLLAR TREE, INC., 7-ELEVEN, INC., FAMILY DOLLAR STORES, INC., THE KROGER CO., DOLLAR TREE STORES, INC., JOHNSON & JOHNSON CONSUMER INC., BIG LOTS, GIANT FOOD LLC, ALBERTSON'S, HARRIS TEETER LLC, DOLGENCORP, LLC,

Defendants-Appellees.

MICHELLE PHIPPEN, INDIVIDUALLY AND AS GENERAL GUARDIAN OF P.P. AND L.A., MINORS, ALISHA DAY, INDIVIDUALLY AND AS MOTHER, GENERAL GUARDIAN OF A.D., A MINOR, SARAH STOKES, INDIVIDUALLY AND AS GENERAL GUARDIAN OF K.G., A MINOR, JUAN EMANUEL BORDOY, INDIVIDUALLY, MARY ELIN ARCE, INDIVIDUALLY AND AS MOTHER OF JUAN EMANUEL BORDOY, AMANDA TRIGLOFF, INDIVIDUALLY AND AS GENERAL GUARDIAN OF R.S., A MINOR, DEANDRE BARBEE, INDIVIDUALLY, JANTAIL BARBEE, INDIVIDUALLY AND AS MOTHER OF DEANDRE BARBEE, LAURIE

COURINGTON, HUNTER COURINGTON, JENNIFER MORROW, ALENA MORROW, CALLISTA BASSETT, ANDREW BASSETT, SONNITA ROBY, INDIVIDUALLY AND AS GENERAL GUARDIAN OF D.P., A MINOR, SHANNON MIKUSKI, BRAYDON MCKENZIE, HEATHER GILLIAM, COLBY GILLIAM, TAYLOR BROWN, TARYNE BURKE, INDIVIDUALLY AND AS MOTHER WITH COURT-APPOINTED GUARDIAN OF ASHTON BURKE, SAMARI SIMS, INDIVIDUALLY, DENISA CULLOM, INDIVIDUALLY AND AS MOTHER OF SAMARI SIMS, COLLIN STOVER, INDIVIDUALLY, DANA STEWART, INDIVIDUALLY AND AS MOTHER OF COLLIN STOVER, RIAN CZAR JOHNSON, INDIVIDUALLY, SHILO RICH, INDIVIDUALLY AND AS MOTHER OF RIAN CZAR JOHNSON, SHEENA SCHNEPP, INDIVIDUALLY AND AS MOTHER AND NATURAL GUARDIAN OF H.S., A MINOR, ZAYNE COSTELLO, INDIVIDUALLY, CRYSTAL ALEXANDER, COURTNEY TILLOTSON, MICHELLE BROWN,

Plaintiffs-Appellants,

v.

WALGREEN CO., JOHNSON & JOHNSON CONSUMER INC., WALMART INC.,

Defendants-Appellees.[*]

APPEAL FROM THE UNITED STATES DISTRICT COURT
FOR THE SOUTHERN DISTRICT OF NEW YORK

Before: CALABRESI, LYNCH, and LEE, *Circuit Judges*.

Plaintiffs-Appellants appeal from judgments of the U.S. District Court for the Southern District of New York (Denise L. Cote, District Judge), dismissing Plaintiffs-Appellants' complaints alleging that Defendants-Appellees failed to

[*] The Clerk of Court is respectfully directed to amend the official caption in this case to conform with the caption above.

warn them that prenatal ingestion of their acetaminophen products could cause attention-deficit/hyperactivity disorder and autism spectrum disorder. In *Rutledge*, the District Court excluded the general causation evidence offered by Plaintiffs-Appellants' five experts, Drs. Baccarelli, Hollander, Pearson, Cabrera, and Louie. In *Phippen*, it excluded the evidence offered by an additional expert, Dr. Ness. The District Court granted summary judgment for Defendants-Appellees in both cases.

With respect to expert testimony, the district court serves a "gatekeeping" function—it is charged with "the task of ensuring that an expert's testimony both rests on a reliable foundation and is relevant to the task at hand." *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 597 (1993). Where a particular technique or theory has gained general acceptance in the scientific community, and an expert reliably applies that methodology to the subject of inquiry, testimony is admissible.

We conclude that the District Court exceeded its discretion by excluding the expert testimony of Drs. Baccarelli, Hollander, and Pearson, but was within its discretion in excluding the testimony of Drs. Cabrera and Louie. In *Phippen*, we conclude that reconsideration of the exclusion of Dr. Ness's testimony is warranted in light of our opinion in *Rutledge*. We further conclude that the District Court correctly declined to dismiss the case on the ground that Plaintiffs-Appellants' failure-to-warn claims were preempted by federal drug labeling laws.

We therefore VACATE and REMAND for further proceedings consistent with this opinion.

ASHLEY C. KELLER, Keller Postman LLC, Chicago, IL; *with* Ashley L.F. Barriere, Keller Postman LLC, Chicago, IL; John J. Snidow & Roseann R. Romano, Keller Postman LLC, Washington, DC, *for Plaintiffs-Appellants Tiffany Rutledge Kristopher White, Bridget McConnell, Alexander Holland, Christine Holland* in Rutledge v. Walgreen Co., *and for Plaintiffs-Appellants Sonnita Roby, Taylor Brown, Michelle Phippen, Amanda Trigloff, Laurie Courington, Hunter Courington, Jennifer Morrow, Alena Morrow, Callista Bassett, Andrew Bassett, Shannon Mikuski, Braydon McKenzie, Heather Gilliam, Colby Gilliam, and Michelle Brown* in Phippen v. Walgreen Co.

Daniel C. Burke, Bernstein Liebhard LLP, New York, NY, *for Plaintiffs-Appellants Juan Emanuel Bordoy, Mary Elin Arce, Deandre Barbee, Jantail Barbee, Taryne Burke, Samari Sims, Denisa Cullom, Collin Stover, Dana Stewart, Rian Czar Johnson, Shilo Rich, Sheena Schnepp, Zayne Costello, Crystal Alexander, and Courtney Tillotson*.

Lindsey Scarcello, Wagstaff & Cartmell, LLP, Kansas City, MO, *for Plaintiff-Appellant Alisha Day*.

JAY P. LEFKOWITZ, Kirkland & Ellis LLP, New York, NY; *with* Cole T. Carter, Kirkland & Ellis, LLP, Chicago, IL, *for Defendant-Appellee Johnson & Johnson Consumer Inc.*

Jeffrey S. Bucholtz & Amy R. Upshaw, King & Spalding LLP, Washington, DC; Matthew Noller, King & Spalding LLP, San Francisco, CA *for Defendants-Appellees Walmart Inc. and Sam's West, Inc*.

Kristen L. Richer, Barnes & Thornburg LLP, Los Angeles, CA, *for Defendants-Appellees CVS Pharmacy, Inc., Walgreen Co., and Costco Wholesale Corporation*.

Amanda Groves, Winston & Strawn LLP, Los Angeles, CA, *for Defendants-Appellees Safeway, Inc., and Albertsons Companies, Inc*.

Joseph A. Lara, Stone | Dean LLP, Woodland Hills, CA, *for Defendant-Appellee The Kroger Company.*

Lori B. Leskin & Mitchell Russell Stern, Arnold & Porter Kaye Scholer LLP, New York, NY; William C. Perdue & Anthony J. Franze, Arnold & Porter Kaye Scholer LLP, Washington, DC, *for Defendants-Appellees 7-Eleven, Inc., Dollar Tree Stores, Inc., and Family Dollar Stores, LLC.*

Anne A. Gruner, Duane Morris LLP, Philadelphia, PA, *for Defendant-Appellee Dolgencorp, LLC*.

Deanne E. Maynard, Morrison & Foerster LLP, Washington, DC; Julie Y. Park & Alexandra Preece Barlow, Morrison & Foerster LLP, San Diego, CA; Alexandra M. Avvocato,

Morrison & Foerster LLP, New York, NY, *for Defendant-Appellee Target Corporation*.

Jeffrey R. White, American Association for Justice, Washington DC, *for Amicus Curiae American Association for Justice, in support of Plaintiffs-Appellants*.

Lawrence P. Eagel, J. Brandon Walker, and Marion C. Passmore, Bragar Eagel & Squire, P.C., New York, NY, *for Amici Curiae Law Professors Anne Bloom, Erwin Chemerinsky, Valerie P. Hans, and Richard L. Jolly, in support of Plaintiffs-Appellants*.

John R. Byrne, Maderal Byrne & Furst PLLC, Coral Gables, FL, *for Amici Curiae Epidemiologists Yinong Young-Xu, Marc Weisskopf, and Graham Colditz, in support of Plaintiffs-Appellants*.

Jennifer B. Dickey & Mariel A. Brookins, Chamber of Commerce of the United States of America, Washington, DC; Joshua J. Fougere & Madeleine Joseph, Sidley Austin LLP, Washington, DC, *for Amicus Curiae Chamber of Commerce of the United States, in support of Defendants-Appellees.*

Raffi Melkonian, Wright, Close & Barger LLP, Houston, TX, *for Amicus Curiae Lawyers for Civil Justice, in support of Defendants-Appellees*.

CALABRESI, *Circuit Judge*:

Plaintiffs-Appellants in these tandem cases bring failure-to-warn claims under state law relating to acetaminophen, the active ingredient in Tylenol and its generic equivalents. They are children, parents, and guardians who allege that prenatal ingestion of acetaminophen caused them or their children to develop attention-deficit/hyperactivity disorder ("ADHD") and/or autism spectrum disorder ("ASD"). Defendants-Appellees are the pharmaceutical companies,

pharmacies, and retailers involved in the manufacturing, marketing, and/or sale of acetaminophen products.

In the cases underlying the first appeal, *Rutledge v. Walgreen Co.*, Plaintiffs-Appellants relied on the general causation testimony of five experts—Dr. Andrea Baccarelli, M.D., Ph.D., Dr. Eric Hollander, M.D., Dr. Brandon Pearson, Ph.D., Dr. Robert Cabrera, Ph.D., and Dr. Stan Louie, Pharm.D.—who opined as to a possible causal relationship between prenatal acetaminophen use and ADHD and ASD. The district court excluded the testimony of those five experts and granted summary judgment to Defendants-Appellees.

The second appeal, *Phippen v. Walgreen Co.*, involves a separate group of Plaintiffs-Appellants alleging injury solely from ADHD. After the district court's initial ruling, those Plaintiffs-Appellants introduced an additional expert, Dr. Roberta Ness, M.D., M.P.H., to testify as to a possible causal relationship between prenatal acetaminophen use and ADHD only. The district court excluded her testimony and granted summary judgment to Defendants-Appellees. We consider these appeals together because of the substantial overlap of the issues they present.

These appeals concern what qualifies as admissible epidemiological testimony in support of a general causal relationship. They arise against the backdrop of significant debate in the relevant scientific communities. That debate has also become political. But the issue before us is not political. It is not about positions taken by elected officials or political appointees. Nor do these appeals require us to decide whether acetaminophen use during pregnancy has adverse effects on fetal development. The issues before us concern the rules of evidence,

specifically the requirements for the admissibility of expert testimony. And so it is important that before we begin, we make clear what we are deciding and what we are not.

We are not deciding whether there is a general causal relationship between acetaminophen and ADHD and/or ASD. We are also not deciding whether the manufacturers of acetaminophen must warn consumers about any alleged risk posed by such a potential causal relationship. And we are certainly not deciding the approach that policymakers concerned with protecting public health should take to regulating the use of acetaminophen. Rather, we are called upon to decide how closely a trial court may scrutinize a qualified expert's conclusions when that expert follows methodologies that are generally accepted in their field, and the standard of reliability required for the admission of expert testimony on issues that are the subject of ongoing scientific debate.

We conclude that, in *Rutledge v. Walgreen Co.*, the district court exceeded its discretion in excluding the expert testimony of Drs. Baccarelli, Hollander, and Pearson. Those concededly qualified experts offered opinions that comport with methodologies applied by other scientists in their fields, and constitute acceptable interpretations of scientific evidence where scientists may, and in fact do, disagree on the ultimate answer to the causal question that they are assessing.

The district court did not, however, abuse its discretion in excluding the testimony of Drs. Cabrera and Louie. The district court was entitled to conclude that Cabrera's opinion was unreliable because it did not weigh or properly synthesize the factors under the so-called Bradford Hill methodology that he

employed. And the district court was similarly entitled to conclude that Louie's opinion did not reliably address the dose and duration exposure threshold for risk of ADHD and ASD because he failed to explain how he extrapolated from the studies he analyzed to reach his final conclusion. We therefore vacate and remand the district court's judgment in *Rutledge*.

In *Phippen*, Plaintiffs-Appellants offered Ness as an expert only after the district court's exclusion of Baccarelli, Hollander, Pearson, Cabrera, and Louie. With the admission of expert testimony from Baccarelli, Hollander, and Pearson, or for other reasons discussed below, it may be that Plaintiffs-Appellants would no longer seek to offer Ness's testimony. We therefore vacate the district court's judgment in *Phippen*, and remand for such further proceedings as the district court finds appropriate, consistent with this opinion.

Finally, Defendants-Appellees argue that in both *Rutledge* and *Phippen* we should affirm the district court's judgment on the alternative ground that federal drug labeling law preempts Plaintiffs-Appellants' failure-to-warn claims. The district court rejected that argument at the pleading stage. Here, the district court did not err. Federal regulations require acetaminophen manufacturers to display verbatim a general pregnancy warning, but they do not prohibit supplemental, specific warnings against plausible risks.

I. BACKGROUND

A. Acetaminophen, ADHD, and ASD

Plaintiffs-Appellants are parents, guardians, and children who claim that they or their children developed ADHD and/or ASD due to prenatal acetaminophen exposure.

Acetaminophen, also called paracetamol or APAP, is the active ingredient in Tylenol and other over-the-counter pain relievers. It is one of the few drugs indicated for use by pregnant women for pain and fever which, if left untreated, can harm both the woman and fetus. It is also a drug intended for systemic absorption (i.e., absorption through the blood stream). As such, when taken by pregnant women, acetaminophen can cross the placental barrier and enter fetal circulation. The FDA requires all such drugs to bear a general warning to pregnant and nursing women advising that they consult "a health professional before use." 21 C.F.R. § 201.63. But, as relevant to Plaintiffs-Appellants' claims, the FDA does not require that acetaminophen carry any warning related to ADHD and/or ASD, nor do Defendants-Appellees offer such a warning.

ADHD and ASD are neurological developmental disorders ("NDDs"). ADHD is characterized by "a persistent pattern of" inattention, hyperactivity, and impulsivity "that interferes with functioning or development." American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders* (5th ed., Text Revision, 2022) ("DSM") at 70. ASD is characterized by a "persistent impairment" in "social communication" and "restricted, repetitive patterns of behavior, interests, or activities." *Id.* at 60. Although both disorders are highly

heritable, a fact that indicates some genetic influence, their precise causes are unknown. *Id.* at 64 (ASD), 71 (ADHD). For example, while some ASD cases are associated with a known genetic mutation, "[a] variety of risk factors for neurodevelopmental disorders, such as advanced parental age, extreme prematurity, or in utero exposures to certain drugs or teratogens like valproic acid, may broadly contribute to risk of [ASD]." *Id.* at 64.

Scientists have been investigating a potential causal relationship between prenatal acetaminophen use and neurodevelopmental disorders for decades but no study has established such a relationship. The FDA opened a Tracked Safety Issue for prenatal acetaminophen exposure in 2014 and has since conducted periodic reviews of the evidence scientists have collected. In each such review, the FDA has noted that while prenatal acetaminophen exposure is associated in some studies with adverse neurological outcomes, study limitations and inconsistent results between studies have prevented the agency from determining causality. *See, e.g.*, Abraham et al., *Functional Neurobehavioral Outcomes and Urogenital Outcomes Associated with Prenatal Acetaminophen Exposure*, U.S. Food and Drug Administration (July 15, 2022), at 33; Abraham et al., *Updated Literature Review of Studies that Examine the Association between Acetaminophen Exposure During Pregnancy and Neurobehavioral or Urogenital Outcomes*, U.S. Food and Drug Administration (March 10, 2023), at 17-18.

In 2021, a group of thirteen authors and seventy-eight signees—consisting of scientists, clinicians, and public health professionals—published a "Consensus Statement" reviewing the literature on prenatal acetaminophen use and ADHD

and ASD.[2] The Consensus Statement concluded that "the combined weight of animal and human scientific evidence is strong enough for pregnant women to be cautioned by health professionals against its indiscriminate use." Bauer et al., *Consensus Statement: Paracetamol Use During Pregnancy—A Call for Precautionary Action*, 17 Nature Revs. Endocrinology 757, 764 (2021) ("Consensus Statement"). And it recommended that acetaminophen "be used by pregnant women cautiously at the lowest effective dose for the shortest possible time." *Id.* But it notably did not conclude that the available data allowed for an inference that a causal relationship exists.[3] *See id.* The Consensus Statement built on and echoed the findings of other scientists who viewed the research as potentially suggesting a causal relationship. *See, e.g.*, Olson & Liew, *Fetal Programming of Mental Health by Acetaminophen? Response to the SMFM Statement: Prenatal Acetaminophen Use and ADHD*, 16 Expert Op. on Drug Safety 1395 (2017) (summarizing research to-date as "increas[ing] the probability that the association is causal"); Gou et al., *Association of Maternal Prenatal Acetaminophen Use with the Risk of Attention Deficit/Hyperactivity Disorder in Offspring: A Meta-Analysis*, 53 Austl. & N.Z. J. of Psychiatry 195 (2019) (similarly concluding that recent research "lend[s] weight to the hypothesis that the association is causal"). And other authorities have noted

[2] To be clear, the "consensus" reflects the views held in common by the paper's signatories. The term does not suggest that the paper reflects a consensus of all experts in the relevant fields of study; as will appear below, it most certainly does not. We use the term because it has become common in the scientific literature as a short-hand reference for the paper in question.

[3] As explained in greater detail below, "[e]pidemiologic methods cannot deductively prove causation"; rather, "epidemiologic evidence can justify an inference, and sometimes a very strong inference, that an agent causes a disease." Federal Judiciary Center, Reference Manual on Scientific Evidence (4th ed. 2025) (RMSE) at 902 n.10.

the potential risk, including the Briggs reference guide on *Drugs in Pregnancy and Lactation* (12th ed. 2022) which acknowledged that prior assessments of acetaminophen as "not . . . caus[ing] embryo-fetal harm . . . must change because of recent data."

The Consensus Statement prompted replies and counterstatements from other scientists and medical bodies. Those publications noted the limitations of the various studies upon which the Consensus Statement relied and expressed concern that the Consensus Statement could lead pregnant women experiencing pain and fever to avoid acetaminophen altogether or to turn to less safe alternatives. *See, e.g.*, Alwan et al., *Paracetamol Use in Pregnancy—Caution Over Causal Inference from Available Data*, 18 Nature Revs. Endocrinology 190 (2022) ("urg[ing] against recommending [] precautionary measures for [acetaminophen] use in pregnancy and against the dissemination of information based on inconclusive and insufficient evidence"); O'Sullivan et al., *Paracetamol Use in Pregnancy—Neglecting Context Promotes Misinterpretation*, 18 Nat. Rev. Endocrinology 385 (2022) (expressing concern that "[t]he overarching societal message that has been drawn from [the] Consensus Statement is that APAP use in pregnancy is unsafe and should be restricted in both use and access"); American College of Gynecologists, *ACOG Response to Consensus Statement on Paracetamol Use During Pregnancy* (Sept. 29, 2021) (noting that "ACOG's clinical guidance remains the same and physicians should not change clinical practice until definitive prospective research is done").

B. Epidemiology and Causation

Epidemiology is the study of the causes, incidence, and distribution of diseases. RMSE at 972. Since ethical constraints render impossible many experiments that might be most probative of a causal relationship (e.g., random clinical trials), epidemiologists often rely instead upon observational studies. In an observational study, epidemiologists "'observe' a group of individuals who have been exposed to an agent of interest, such as . . . an industrial chemical, and compare them with another group of individuals who have not been exposed." RMSE at 906. An observational study thus allows epidemiologists to determine whether an association exists between an agent and a health outcome. But an association "does not necessarily mean that there is a cause-effect relation" between the agent and that outcome. RMSE at 921. Accordingly, "[t]o make a judgment about causation," epidemiologists consider observational studies identifying an association through the lens of "several key inquiries." RMSE at 971, 973. One method generally accepted by the courts for structuring that process is consideration of the Bradford Hill criteria.[4] *See, e.g.*, *Sarkees v. E.I. Dupont De Nemours & Co.*, 15 F.4th 584, 591–92 (2d Cir. 2021); *In re Zoloft (Setraline Hydrochloride) Prods. Liab. Litig.*, 858 F.3d 787, 796 (3d Cir. 2017).

The Bradford Hill criteria are a set of factors that epidemiologists examine to assess whether an observed association is causal in nature. RMSE at 973. No single Bradford Hill factor is required to infer causation. Nor are the criteria "an

[4] The Bradford Hill method takes its name from a 1965 paper by the epidemiologist Sir Austin Bradford Hill describing the approach. *See* RMSE at 973.

exhaustive []or a necessary list." *Zoloft*, 858 F.3d at 796. In a Bradford Hill analysis, an epidemiologist typically identifies the body of published scientific literature that is relevant to their question and then analyzes the results of each of those studies for the presence, or lack thereof, of each of the Bradford Hill criteria. The criteria as presented in the Federal Judiciary Center's Reference Manual on Scientific Evidence[5] are:

(1) Replication of the findings (also referred to as "consistency"): When the outcomes of a study are observed "in different populations and by different investigators," this supports a causal determination. RMSE at 981.

(2) Strength: "Larger relative risks (or stronger associations using other statistical measures) are often believed to be more likely to be causal than smaller ones." RMSE at 977.

(3) Specificity: Where an "exposure is associated only with a single disease or type of disease," this may be strong evidence in support of causality. RMSE at 984.

(4) Dose-response relationship: If an exposure to a risk factor causes a disease, then "higher exposures would generally be expected to increase the incidence or severity of that disease." RMSE at 977. Dose responses may take different shapes, including a straightforward linear relationship, a linear relationship after the dose exceeds a particular

[5] Epidemiologists have formulated the Bradford Hill criteria in different ways, for example, referring to the same criterion under a different label or consolidating multiple criteria under a single one.

threshold, or a non-monotonic effect such as when both under-exposure and over-exposure to a particular agent causes impaired outcomes. RMSE at 978–81.

(5) Temporal relationship: Exposure to the risk factor must precede the disease, because "[i]f the exposure occurs after the disease develops, it cannot have caused the disease." RMSE at 975.

(6) Biological plausibility: When a proposed causal relationship is consistent with "current biological knowledge," the causal inference is strengthened. RMSE at 983. This is "not an easy criterion to use and depends upon existing knowledge about the mechanisms by which the disease develops." RMSE at 982. "The mechanisms of some diseases are understood quite well based on evidence . . . whereas other mechanism explanations are merely hypothesized—although hypotheses are sometimes also accepted under this factor." RMSE at 983.

(7) Consistency with other relevant knowledge (also referred to as "coherence"): A causal relationship should be consistent with other information known about the disease. RMSE at 985.

(8) Cessation of exposure (also referred to as "experiment"): When experimental evidence shows that "eliminating exposure reduces the incidence of disease," this supports a causal relationship. RMSE at 984.

In conducting this inquiry, an epidemiologist must also take care to consider "the possibility that an observed association is caused by something other than the exposure under study," such as "bias and confounding." RMSE at 984.

C. Statutory and Regulatory Framework

"Under the Federal Food, Drug, and Cosmetic Act, 21 U.S.C. § 301 *et seq.* ('FFDCA'), a new drug may not enter interstate commerce unless [the] FDA determines that it is generally recognized as safe and effective ('GRAS/E') for the particular use described in its product labeling." *NRDC v. FDA*, 710 F.3d 71, 75 (2d Cir. 2013). Two such pathways to a GRAS/E determination are relevant here: the New Drug Application ("NDA") process and the monograph system.

Under the NDA process, "a manufacturer seeking federal approval to market a new drug must prove that it is safe and effective and that the proposed label is accurate and adequate." *PLIVA, Inc. v. Mensing*, 564 U.S. 604, 612 (2011). "The FDA's premarket approval of a new drug application includes the approval of the exact text in the proposed label." *Wyeth v. Levine*, 555 U.S. 555, 568 (2009). After a manufacturer receives such approval, it remains charged with "ensuring that its warnings remain adequate as long as the drug is on the market." *Id.* at 571. Hence a manufacturer may make post-approval label changes that "add or strengthen a contraindication, warning, precaution, or adverse reaction" based on "newly acquired information." 21 C.F.R. § 314.70(c)(6)(iii).

Over-the-counter ("OTC") drugs, however, may also be approved through the monograph system. *See* 21 C.F.R. § 330.10; 21 U.S.C. § 355. "Under this system, FDA issues a detailed regulation—a 'monograph'—for each therapeutic class of OTC drug products." *NRDC*, 710 F.3d at 75. Each final monograph "establish[es] conditions under which a category of OTC drugs . . . are generally recognized as safe and effective and not misbranded." 21 C.F.R. § 330.10(a)(9). A final

monograph "may include any conditions relating to active ingredients, labeling indications, warnings and adequate directions for use . . . necessary and appropriate for the safety and effectiveness of drugs covered by the monograph," *id.* § 330.10(a)(5)(i), but need not dictate the full label that manufacturers must apply to a product marketed under it. Moreover, manufacturers marketing a specific product under the monograph "are not required to submit labeling to the agency for preapproval." Over-The-Counter Human Drugs; Labeling Requirements, 64 Fed. Reg. 13254, 13271 (Mar. 17, 1999).

Over-the-counter acetaminophen products are marketed pursuant to the monograph for Internal Analgesic, Antipyretic, and Antirheumatic Drug Products ("IAAA"). *See* U.S. Food and Drug Administration, Over-the Counter (OTC) Monograph M013: Internal Analgesic, Antipyretic, and Antirheumatic Drug Products for Over-the Counter Human Use (Oct. 14, 2022).[6] The IAAA monograph sets forth various labeling requirements. For example, acetaminophen labels must specify only the indications for use established in the monograph. The monograph does not, however, expressly prohibit additional warnings not specified therein.

Whether approved through the NDA or monograph system, OTC drugs that are intended for systemic absorption must also contain a general pregnancy and breast-feeding warning. 21 C.F.R. § 201.63 ("Pregnancy Warning

[6] While the IAAA monograph was not finalized until 2022, it was first proposed as a tentative final monograph in 1988, during which time it had the status of a proposed rule. *See* Internal Analgesic, Antipyretic, and Antirheumatic Drug Products for Over-the-Counter Human Use; Tentative Final Monograph, 53 Fed. Reg. 46204 (Nov. 16, 1988).

Regulation"). Acetaminophen is one such drug. The Pregnancy Warning Regulation states:

> The labeling for all over-the-counter (OTC) drug products that are intended for systemic absorption, unless specifically exempted, shall contain a general warning under the heading "Warning" (or "Warnings" if it appears with additional warning statements) as follows: "If pregnant or breast-feeding, ask a health professional before use." [first four words of this statement in bold type] In addition to the written warning, a symbol that conveys the intent of the warning may be used in labeling.

Id. § 201.63(a) (bracketed text in original). Because the warning language identified in the Pregnancy Warning Regulation is "established and identified by quotation marks," a separate regulation, the Exact Language Regulation, requires that the warning appear in the "exact language" specified. *Id.* § 330.1(c)(2). The Pregnancy Warning Regulation further provides, however, that

> [w]here a specific warning relating to use during pregnancy or while nursing has been established for a particular drug product in [an NDA] or for a product covered by an OTC drug final monograph [then such] specific warning shall be used in place of the warning in paragraph (a) of this section ["If pregnant or breast-feeding, ask a health professional before use"], unless otherwise stated in the NDA or in the final OTC drug monograph.

Id. § 201.63(b).

D. Procedural Background

Several months after publication of the Consensus Statement that suggested possible links between acetaminophen and ADHD and ASD, Plaintiffs-Appellants began to file suits against the manufacturers and retailers of acetaminophen,

alleging that prenatal exposure to the drug caused them or their children to develop ADHD and/or ASD. Those cases were consolidated and transferred to the Southern District of New York by the Judicial Panel on Multidistrict Litigation pursuant to 28 U.S.C. § 1407.

1. Preemption Defense

Two Defendants-Appellees—Walmart, Inc. and Johnson & Johnson Consumer Inc.—moved to dismiss three failure-to-warn actions on the ground that federal law preempted the addition of ADHD and/or ASD warnings to pregnant women on acetaminophen labels. The district court denied those motions in separate decisions, concluding that while federal law required acetaminophen manufacturers to display a general pregnancy warning, that requirement did not preclude manufacturers from including an additional warning specific to the risk of ADHD and/or ASD. *In re Acetaminophen—ASD-ADHD Prods. Liab. Litig.*, No. 22-md-3043, 2022 WL 17348351 (S.D.N.Y. Nov. 14, 2022); *In re Acetaminophen—ASD-ADHD Prods. Liab. Litig.*, No. 22-md-3043, 2023 WL 3026412 (S.D.N.Y. Apr. 20, 2023).

2. Motions to Exclude Expert Testimony under Rule 702 and for Summary Judgment

Following consolidation, the parties agreed first to conduct discovery related to general causation (i.e., whether a causal relationship generally exists between prenatal acetaminophen exposure and ADHD and ASD) and, then, if Plaintiffs-Appellants' experts survived Rule 702 motions, to proceed with the remainder of discovery, including specific causation (i.e., whether a particular

plaintiff developed ADHD and/or ASD because of prenatal acetaminophen exposure). While Plaintiffs-Appellants bring their claims under the tort law of various states, there is no dispute that "all fifty states require some evidence of general causation in products liability cases involving complex products liability (or medical) issues." *In re Mirena IUS Levonorgestrel-Related Prods. Liab. Litig. (No. II)*, 982 F.3d 113, 124 (2d Cir. 2020) (internal quotation marks omitted). Plaintiffs-Appellants produced five experts to opine on general causation.

Dr. Baccarelli is an epidemiologist and physician, an elected member of the National Academy of Medicine, and dean of the Harvard T.H. Chan School of Public Health. App'x 5417. He identified and analyzed the peer-reviewed literature on prenatal acetaminophen exposure, then applied two epidemiological methodologies—the Bradford Hill criteria and the Navigation Guide[7]—to synthesize that evidence. Under each methodology, Baccarelli concluded that the evidence supports a finding of causality.

Dr. Hollander is a psychiatrist and professor at the Albert Einstein College of Medicine. He opined as to the interconnectedness of neurodevelopmental disorders like ADHD and ASD, explaining that it is appropriate for scientists to consider the two outcomes together, as well as to consider studies with symptomatic endpoints, when assessing whether a causal relationship exists

[7] The Navigation Guide is a method distinct from the Bradford Hill analysis. According to Baccarelli, the Navigation Guide is used "to more readily evaluate causal relationships for toxic and environmental harms." App'x 1745. It involves the "systematic rating and review of" individualized studies addressing a potential causal relationship "for bias, strength of evidence, and other indicia of study quality." *Id.*

between acetaminophen and ADHD and ASD. App'x 2475. Plaintiffs-Appellants offered Hollander's testimony as support for Baccarelli's Bradford Hill analysis, which considered ADHD and ASD together, and for Baccarelli's decision to consider studies that utilized symptomatic endpoints, in addition to those that relied upon ADHD and/or ASD as diagnostic outcomes. In his rebuttal report, Hollander also provided his own Bradford Hill analysis.

Dr. Pearson is a toxicologist at Columbia University. He specializes in preclinical research relating to neurodevelopmental disorders.[8] Pearson's report identified and explained the biological mechanisms by which acetaminophen exposure could cause ADHD and ASD based on the preclinical literature. App'x 2021.

Dr. Cabrera is a teratologist and geneticist at Baylor College of Medicine. Teratology is the study of "abnormalities, malformations, and developmental disorders that occur during prenatal development." App'x 2207. Cabrera reviewed the available literature and used two established methodologies—weight-of-the-evidence[9] and the Bradford Hill criteria—to conclude that acetaminophen can cause ADHD and ASD. He also asserted the biological mechanisms through which acetaminophen exposure can be a cause of ADHD and

[8] Preclinical studies involve animal and other non-human testing.

[9] As its name suggests, under the weight-of-the-evidence approach, an epidemiologist synthesizes a review of the available scientific literature—including "clinical observations, case reports, epidemiological and animal studies, toxicological experiments, [and] exposure data," alongside "the internal and external validity of th[at] data"—to arrive at an opinion on causation. App'x 2216 n.4.

ASD by affecting neurodevelopment, using the Adverse Outcome Pathway framework.

Dr. Louie is a professor of clinical pharmacy at the University of Southern California. He reviewed the available literature and opined on plausible biological mechanisms through which acetaminophen can cause ADHD and ASD, and the dose and duration at which such effects could take place. He opined that the risk of ADHD and ASD increases when acetaminophen is taken for at least 28 cumulative days during pregnancy. App'x 2737.

The Defendants-Appellees called six experts, only one of whom, Dr. Stephen Faraone, is directly relevant to this appeal. Faraone is a professor of psychiatry and neuroscience at the State University of New York Upstate Medical University. In 2021, he coordinated publication of a separate consensus statement that compiled "findings with [a] strong evidence base" regarding ADHD. *See* Faraone et al., *The World Federation of ADHD International Consensus Statement: 208 Evidence-Based Conclusions about the Disorder*, 128 Neuroscience & Biobehavioral Revs. 789 (2021). One finding recognized by that paper was a correlation between prenatal acetaminophen use and ADHD in children. *See id.* at 795. In this case, however, Faraone testified for the Defendants-Appellees that the scientific literature does not support a causal inference between prenatal acetaminophen and ADHD.

Following presentation of the expert reports, rebuttal reports, and expert depositions, the parties each submitted Rule 702 motions. After oral argument, the district court granted Defendants-Appellees' Rule 702 motion in its entirety,

excluding the testimony of Baccarelli, Hollander, Pearson, Cabrera, and Louie, and denied Plaintiffs-Appellants' Rule 702 motion as moot. The district court reasoned that the testimony of each expert did not reflect a reliable application of epidemiological methodologies to the question at hand, and granted summary judgment to Defendant-Appellees in approximately 550 cases in the MDL. Plaintiffs-Appellants appealed that decision in *Rutledge.*

A separate group of Plaintiffs-Appellants who had filed their claims after the district court issued its initial Rule 702 decision subsequently sought to introduce their own expert to testify on general causation for acetaminophen and ADHD *only*. That expert, Dr. Ness, is an epidemiologist who previously served as dean of the University of Texas School of Public Health and as a professor in public health at the University of Texas Houston until her retirement. She conducted a Bradford Hill analysis that was responsive to the district court's analysis and exclusion of the previous five experts, and testified that there is a causal relationship between prenatal acetaminophen exposure and ADHD.

Defendants-Appellees filed a Rule 702 motion to exclude Ness. The district court granted the motion, holding that Ness's testimony did not appropriately account for genetic confounding and improperly cherry-picked among studies. Like their predecessors, this subset of Plaintiffs-Appellants was ordered to show cause why final judgment should not be entered. They contested the exclusion of Ness and argued, in the alternative, that summary judgment was not appropriate because Faraone's earlier statements in support of a causal connection between acetaminophen and ADHD could allow a reasonable jury to find in Plaintiffs-

Appellants' favor on the issue of general causation. The district court granted summary judgment for Defendants-Appellees, continuing to exclude Ness and explaining that "even if each [Faraone] statement were admissible, the statements would not constitute reliable evidence of general causation in support of the plaintiffs' theory." *In re Acetaminophen – ASD-ADHD Prods. Liab. Litig.*, 2024 WL 3874183, at *4 n.6 (S.D.N.Y. Aug. 20, 2024). That judgment is the basis of the appeal in *Phippen*.

Plaintiffs-Appellants now bring these tandem appeals challenging, in *Rutledge*, the district court's exclusion of the five general causation experts for ADHD and ASD (Baccarelli, Hollander, Pearson, Cabrera, and Louie), and in *Phippen*, one general causation expert as to ADHD (Ness). The *Phippen* Plaintiffs-Appellants further argue in the alternative that, even if all experts remain excluded, summary judgment was not appropriate because a reasonable jury could rely on Faraone's statements to find for Plaintiffs-Appellants. Defendants-Appellees contend that these experts were properly excluded and that, even if they weren't, we should affirm the district court's judgments because the FDA's regulations preempt any state law that would require Defendants-Appellees to supplement the FDA's mandated pregnancy warning on their products.

II. Discussion

A. Rule 702 Legal Framework

The admissibility of expert scientific testimony is governed by the Federal Rules of Evidence. Rule 702 provides that:

> [a] witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify . . . if the proponent demonstrates to the court that it is more likely than not that: (a) 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; (b) the testimony is based on sufficient facts or data; (c) the testimony is the product of reliable principles and methods; and (d) the expert's opinion reflects a reliable application of the principles and methods to the facts of the case.

See also Daubert v. Merrell Dow Pharms., 509 U.S. 579, 598 (1993). This standard requires district courts to look carefully at the qualifications and methodology of each expert.

1. Qualifications

"To determine whether a witness qualifies as an expert, courts compare the area in which the witness has superior knowledge, education, experience, or skill with the subject matter of the proffered testimony." *United States v. Tin Yat Chin*, 371 F.3d 31, 40 (2d Cir. 2004). "Experts need not conduct studies of their own in order to opine on a topic; a review of other studies and scientific literature can be enough to qualify experts to testify and to make that proposed testimony reliable." *In re Mirena IUD Prods. Liab. Litig.*, 169 F. Supp. 3d 396, 412 (S.D.N.Y. Mar. 8, 2016), citing *McCullock v. H.B. Fuller Co.*, 61 F.3d 1038, 1042–43 (2d Cir. 1995).

2. Reliability

In addition to confirming that experts are qualified to offer testimony, Rule 702 charges district courts with the "task of ensuring that an expert's testimony both rests on a reliable foundation and is relevant to the task at hand." *Daubert*, 509 U.S. at 597. To determine reliability, the district court may consider whether

the expert's theory or technique can and has been tested; whether it has been subjected to peer review and publication; whether it has a known error rate or standards to control its operation; and its general acceptance in the relevant scientific community. *Id.* at 593–94. The objective of this gatekeeping requirement 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). "In deciding whether a[n] . . . expert's analysis is []reliable, the district court should undertake a rigorous examination of the facts on which the expert relies, the method by which the expert draws an opinion from those facts, and how the expert applies the facts and methods to the case at hand." *Amorgianos v. Nat'l R.R. Passenger Corp.*, 303 F.3d 256, 267 (2d Cir. 2002).

While the focus of the court's *Daubert* inquiry should be "on principles and methodology, not on the conclusions that they generate," *Daubert*, 509 U.S. at 595, "conclusions and methodology are not entirely distinct from one another," *Gen. Elec. Co. v. Joiner*, 522 U.S. 136, 146 (1997). Accordingly, "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." *Id.* Consequently, "when an expert opinion is based on data, a methodology, or studies that are simply inadequate to support the conclusions reached, *Daubert* and Rule 702 mandate the exclusion of that unreliable opinion testimony." *Ruggiero v. Warner-Lambert Co.*, 424 F.3d 249, 255 (2d Cir. 2005), quoting *Amorgianos*, 303 F.3d at 266.

As the Rules Committee has explained, "[i]t will often occur that experts come to different conclusions based on contested sets of facts. Where that is so the [rule] does not necessarily require exclusion of either side's experts." Fed. R. Civ. P. 702 committee note to 2023 amendment. That is because a party "do[es] not have to demonstrate to the judge by a preponderance of the evidence that the assessments of their experts are correct, they only have to demonstrate by a preponderance of the evidence that their opinions are reliable." *Id.* (internal quotation marks omitted). "The evidentiary standard of reliability is lower than the merits standard of correctness." *Id.* (internal quotation marks omitted).

3. Standard of Review

"We review a district court's determination to admit or exclude expert testimony under *Daubert* for abuse of discretion." *Amorgianos*, 303 F.3d at 264. A decision to admit or exclude expert scientific testimony is not an abuse of discretion unless it is "manifestly erroneous." *In re Mirena (No. II)*, 982 F.3d at 122 (2d Cir. 2020) (internal quotation marks omitted).

B. Rule 702 Application

As to the first prong of the Rule 702 analysis in *Rutledge*, we agree with the district court that "[e]ach of the parties' experts is eminently qualified." 707 F. Supp. 3d at 317.

On the second prong, while the district court's review of the expert testimony at issue in this case was thorough and well-documented, we conclude that, in the case of Baccarelli, Hollander, and Pearson, the court went beyond its proper role as gatekeeper by excluding experts whose testimony was consistent

with the methodologies actually employed in their field; by substituting its own understanding of various epidemiological criteria for that of the expert scientists; and by faulting the scientists for reaching particular conclusions where there is disagreement in the scientific community over the proper interpretation of the body of evidence. What follows is a non-exhaustive list of the instances in which we hold the district court to have been in error. These instances, we find, are sufficient to vacate the district court's opinion as to Baccarelli, Hollander, and Pearson and to hold their testimony to be admissible. It bears repeating that this is not to say that we deem the opinions of those experts correct on the issue of causation, we hold only that the district court erred in deeming their testimonies unreliable and therefore not worthy of consideration by a jury. And because we find that the district court erred in its assessment in *Rutledge*, we need not reach the district court's assessment of Ness in *Phippen*.

As will be made clear, our analysis primarily focuses on Baccarelli's opinion, which also sits front and center in the parties' briefing and the district court's analysis. That opinion is the broadest in scope and supplies the foundation for Plaintiffs-Appellants' assertion of a causal relationship between prenatal exposure to acetaminophen and ASD and ADHD. The other experts' opinions, while also noteworthy, serve to support and corroborate aspects of Baccarelli's analysis.

1. Baccarelli

a. Transdiagnostic Bradford Hill analysis

The district court first rejected Baccarelli's testimony because of how he structured his Bradford Hill analysis. Baccarelli performed a single Bradford Hill

analysis to examine the effect of prenatal acetaminophen on ADHD, ASD, and neurodevelopmental symptoms consistent with those disorders. The parties and the district court have referred to this as a "multiple-outcomes" or "transdiagnostic" approach. As inputs to that analysis, Baccarelli considered peer-reviewed studies that used both diagnostic and non-diagnostic endpoints. Diagnostic endpoint studies examine populations that have been diagnosed with the condition of interest (e.g., ADHD and/or ASD), whereas non-diagnostic endpoint studies use outcomes such as symptoms or closely related conditions (e.g., neurodevelopmental deficits associated with ADHD and/or ASD). In other words, he performed a single Bradford Hill analysis on both ADHD and ASD together and used studies that examined both symptomatic and diagnostic outcomes.

Baccarelli explained this methodological choice in his report: "[D]ifferent NDDs often have shared/overlapping symptomology as well as shared biological pathways or causes . . . . Therefore, in assessing the association between acetaminophen exposure in utero and neurodevelopmental disorders like ASD and ADHD, it is important to consider not just the clinical diagnoses but also the neurodevelopmental outcomes, including the symptoms." App'x 1778. He acknowledged in his rebuttal report that while "studies showing an association between prenatal [acetaminophen] exposure and clinical *diagnoses* of ADHD and ASD provide *particularly* powerful evidence of a causal relationship," that "in no way suggests that researchers should blind themselves to the results from studies using other endpoints that are obviously relevant to the question at hand." App'x 2892.

The district court held that this methodological choice was inappropriate because, while ADHD and ASD share some symptoms, "the diagnostic criteria of the two disorders are undeniably distinct." 707 F. Supp. 3d at 340. And the district court further concluded that the focus on symptomatic outcomes rendered Baccarelli's testimony irrelevant because "this litigation is brought to obtain recovery on behalf of those who have been diagnosed with ASD or ADHD, not on behalf of anyone with, for example, a deficit in communication or self-regulation." *Id.* at 339**.** The district court also asserted that the multiple-outcomes approach "obscured limitations in the scientific literature," such that "[i]f the studies for either ASD or ADHD were subjected to their own individual Bradford Hill analysis, it would be easier to discern whether there was actual support for a finding that prenatal exposure to acetaminophen causes either ASD or ADHD." *Id.*

That ruling was erroneous because it penalized Baccarelli for using a methodology that epidemiologists routinely use—in other words, for testimony that embodies "the same level of intellectual rigor that characterizes the practice of an expert in the relevant field." *Kumho Tire Co.*, 526 U.S. at 152.

Scientists have conducted multiple-outcome Bradford Hill analyses examining, for example, the relationship between air pollution, mental diseases, and mortality; between e-cigarettes and various types of oral cancers; and between psychological factors and irritable bowel disease (encompassing Crohn's disease

and ulcerative colitis).[10] Moreover, other scientists have, independently, examined the Bradford Hill criteria for ADHD and ASD together. *See* Alemany et al., *Prenatal and Postnatal Exposure to Acetaminophen in Relation to Autism Spectrum and Attention-Deficit and Hyperactive Symptoms in Childhood*, 36 Euro. J. Epidemiology 993, 1000 (2021). And the FDA has likewise considered the relationship between "in utero [acetaminophen] exposure and neurobehavioral and urogenital outcomes." Abraham 2023, at 3. Indeed, in its own weight-of-the-evidence review, Johnson & Johnson examined whether prenatal acetaminophen exposure could cause a broad category of "neurocognitive/neurodevelopmental-related events," including developmental disorders or delays, neurological disorders or delays, ADHD, speech and language disorders, and ASD. App'x 6610, 6613.

Baccarelli's use of studies that examine non-diagnostic endpoints, in addition to those that used diagnostic ones, likewise did not render his testimony unreliable. Other scientists not involved in this litigation studying the connection between prenatal acetaminophen exposure and ADHD and ASD have routinely utilized symptomatic endpoints alone or in combination with diagnostic ones—

[10] *See, e.g.*, John P.A. Ioannidis, *Air Pollution as Cause of Mental Disease: Appraisal of the Evidence*, 17 PLOS Biology (2019) (evaluating the connection between "air pollution" and "mental disease," including depression, bipolar disorder, schizophrenia, and personality disorder); Raj et al., *Reviewing the Oral Carcinogenic Potential of E-Cigarettes Using the Bradford Hill Criteria of Causation*, 9 Translational Cancer Rsch. 3142 (2020) (evaluating "e-cigarettes" and "oral cancer"); Schoultz et al., *Assessment of Causal Link Between Psychological Factors and Symptom Exacerbation in Inflammatory Bowel Disease: A Systematic Review Utilising Bradford Hill Criteria and Meta-Analysis of Prospective Cohort Studies*, 9 Systematic Revs. (2020) (evaluating "irritable bowel disease" and "psychological stress," which encompassed a "variety of minor to major psychological factors").

several of which Baccarelli cites. *See, e.g.*, App'x 1840 (citing Avella-Garcia 2016,[11] which measured "autism spectrum symptoms" and "inattention and hyperactivity/impulsivity symptoms"); App'x 1834 (citing Masarwa 2018,[12] a meta-analysis that discussed studies with both diagnostic and symptomatic endpoints). At least one of Defendants-Appellees' experts also discussed symptom-endpoint studies in a textbook chapter. *See* App'x 4836 (Dr. Alexander Kolevzon citing Masarwa 2018). So did Johnson & Johnson's own scientists in their internal review of the literature on prenatal acetaminophen exposure. *See* App'x 6652 (citing Avella-Garcia 2016); App'x 6655 (citing Bornehag 2018,[13] which measured "language development"); App'x 6649 (citing Thompson 2014,[14] which measured "ADHD symptoms"). In fact, Johnson & Johnson's review of the literature noted Avella-Garcia's use of symptomatic evidence as a *strength* of that study, because it "assessed milder ADHD and Autism Spectrum Condition (ASC) symptoms which are much more prevalent in the population." App'x 6654.

We do not mean to suggest that all Bradford Hill analyses that simultaneously examine multiple outcomes, or that use symptomatic as well as diagnostic endpoints, are reliable and admissible under Rule 702. We can imagine,

---

[11] Avella-Garcia et al., *Acetaminophen Use in Pregnancy and Neurodevelopment: Attention Function and Autism Spectrum Symptoms*, 45(6) Int. J. Epidemiology 1987 (2016).

[12] Masarwa et al., *Prenatal Exposure to Acetaminophen and Risk for Attention Deficit Hyperactivity Disorder and Autistic Spectrum Disorder: A Systematic Review, Meta-Analysis, and Meta-Regression Analysis of Cohort Studies*, 187(8) Am. J. Epidemiology 1817 (2018).

[13] Bornehag et al., *Prenatal Exposure to Acetaminophen and Children's Language Development at 30 Months*, 51 Euro. Psych. 91 (2018).

[14] Thompson et al., *Association Between Acetaminophen Use During Pregnancy and ADHD Symptoms Measured at Ages 7 and 11 Years*, 9(9) PLoS One 1 (2014).

for example, conditions which are sufficiently distinct so that a single, joint, analysis could not be performed in a reliable manner. There may well be conditions for which symptoms alone, without a diagnosis, may well be of little practical use to scientists looking to understand what may cause a disease. But that is not the case here, where the technique at issue has "attract[ed]" "[w]idespread acceptance" in the relevant scientific "community," *Daubert*, 509 U.S. at 594 (internal quotation marks omitted), including in work by both the FDA and Defendants-Appellees' own scientists outside of this litigation.[15]

To restate the obvious, we do not conclude that *we* think the structure Baccarelli employed for his Bradford Hill analysis is the best one, nor do we have any prediction, one way or another, about whether a jury will find it persuasive. Those are not the issues before us. The single, multiple-outcomes analysis may well prove unconvincing to a jury for precisely the reasons that the district court identifies, but that does not render the expert opinion excludable for the purposes of Rule 702.

---

[15] The district court also said the multiple-outcomes analysis should be excluded because it "obscure[s] limitations in the scientific literature" such that, if the analysis were focused on a single disorder, the reliability of such testimony would be easier to determine. 707 F. Supp. 3d at 339. But the concern that the multiple-outcomes analysis prevents a factfinder from disentangling the studies assessing ASD from those assessing ADHD is undermined by the district court's own opinion doing exactly that. While it might have been easier to sort through the Rule 702 motions if Baccarelli's Bradford Hill analysis had treated ADHD and ASD separately, the district court still ably sifted through the various studies, opined on the limitations in the literature regarding ASD, and thoroughly analyzed the issue of genetic confounding.

b. Bradford Hill Factors

The district court next held that, even if it were methodologically reliable for Baccarelli to assess multiple outcomes and to rely on studies that used both diagnostic and symptomatic endpoints, Baccarelli applied that method in an unreliable manner. We find that the court, in so holding, improperly (i) substituted its own definitions of certain Bradford Hill factors for those of other epidemiologists, and (ii) penalized Baccarelli for drawing plausible conclusions well within "the range where experts might reasonably differ." *Kumho Tire*, 526 U.S. at 152. What follows is a non-exhaustive discussion of instances in which the district court overstepped its gatekeeping function.

i. The definition of individual factors

The district court's discussion of Baccarelli's testimony on dose response, biological plausibility, strength, and specificity illustrate the first point. On dose response, the district court faulted Baccarelli for failing to grapple with "a key issue in the underlying studies," which was that "none were able to record the actual dosages taken by pregnant women." 707 F. Supp. 3d at 350 (emphasis in original). But the district court cited no scientific authority for the proposition that dose-response analysis requires data on the precise doses that study participants ingested. And multiple meta-analyses of the same literature that Baccarelli analyzed made dose-response judgments similar to his. *See, e.g.*, Alemany 2021 (recognizing that multiple prior studies found a dose-response relationship); Ricci

2023[16] (finding that the association between acetaminophen exposure and ADHD was strongest in children with the highest duration of exposure, "suggesting a dose-response effect"). Indeed, the district court noted that "[t]he closest approximation to dose *in the underlying studies* is days of use." 707 F. Supp. at 350 (emphasis added). In other words, the district court faulted Baccarelli for analyzing acetaminophen dose response in exactly the manner used by scientists in his field, utilizing the best available proxy for data that would rarely if ever be available in observational studies, and that for ethical reasons could not be collected by a controlled experimental study.

On biological plausibility, the district court discredited Baccarelli's testimony because "[a]t present, the precise physiological process or processes by which these conditions, or NDDs more generally, develop are unknown." *Id.* It found that because "[s]cientists have at best developed hypotheses" as to the biological mechanisms by which acetaminophen causes ADHD and ASD, "Baccarelli's conclusion that this factor is satisfied . . . fail[s] to reflect a reliable application of scientific principles." *Id.*

There is no doubt that any causal relationship between acetaminophen and ADHD and ASD is quite uncertain. But the Bradford Hill methodology does not require certainty regarding the mechanisms by which exposure causes the condition of interest—indeed, if such proof were available, there would be no need for the multi-factor epidemiological inquiry in the first place. The Bradford Hill

[16] Ricci et al., *In Utero Acetaminophen Exposure and Child Neurodevelopmental Outcomes: Systematic Review and Meta-Analysis*, 37 Paediatric Perinatal Epidemiology 473 (2023).

methodology asks no more than whether there is some such "biologically *plausible* mechanism." App'x 1906 (emphasis added).[17] In other words, it "asks whether the hypothesized causal link is credible in light of what is known from science and medicine" about the alleged toxin and its biological effects. *Milward v. Acuity Specialty Prods. Grp., Inc.*, 639 F.3d 11, 25 (1st Cir. 2011). And Baccarelli's approach to biological plausibility, in this respect, finds support in the practices of other epidemiologists outside the courtroom, in assessing both whether acetaminophen exposure causes ADHD and ASD,[18] and in investigations into the potential causes of other conditions.[19] Baccarelli hence applied the biological plausibility criterion with "the same level of intellectual rigor that characterizes the practice of an expert" in his field. *Kumho Tire Co.*, 526 U.S. at 152. Of course, the jury remains free to consider the fact that scientists have not definitively identified a mechanism of causation and to assign the weight to that fact it deems appropriate.

---

[17] *See also* Sir Austin Bradford Hill, *The Environment and Disease: Association or Causation?*, 58 Proc. Royal Soc'y Med. 295, 298 (1965) ("It will be helpful if the causation we suspect is biologically plausible. But this is a feature I am convinced we cannot demand. What is biologically plausible depends upon the biological knowledge of the day.").

[18] *See, e.g.*, Gustavson et al., *Acetaminophen Use During Pregnancy and Offspring Attention Deficit Hyperactivity Disorder -- A Longitudinal Sibling Control Study*, 1 JCPP Advances at 2 (2021) ("Biologically plausible mechanisms for effect on fetal neurodevelopment include oxidative stress and neurotoxicity."); Chen et al., *Prenatal Exposure to Acetaminophen and the Risk of Attention-Deficit/Hyperactivity Disorder: A Nationwide Study in Taiwan*, 80 J. Clin. Psychiatry at 5 (2019) (identifying "[s]everal biologically plausible contentions," including that "acetaminophen may alter the intrauterine immune system and increase the predisposition for oxidative stress and inflammation").

[19] *See, e.g.*, Bello et al., *Does Human Papillomavirus Play a Causative Role in Prostate Cancer? A Systematic Review Using Bradford Hill's Criteria*, 15 Cancers at 7 (2023) (finding biological plausibility satisfied in part based on "the oncogenic role of HPV in other types of cancers"); Davidson & Smith, *The Bradford Hill Criteria and Zinc-Induced Anosmia; A Causality Analysis*, 136(7) Archives of Otolaryngology – Head Neck Surgery 673, 675 (2010) (concluding that biological plausibility is satisfied based on experimental animal evidence showing that zinc exposure causes temporary anosmia).

In its analysis of strength, the district court first noted that most of the studies dealing with ADHD and/or ASD show a risk ratio of between 1.0 and 2.0,[20] which is lower than the ratios present in other cases in this Circuit. *See, e.g.*, *In re Mirena IUS Levonorgestrel-Related Prods. Liab. Litig. (No. II)*, 341 F. Supp. 3d 213, 243 (S.D.N.Y. 2018), *aff'd* 982 F.3d 113 (2d Cir. 2020) (risk ratios of 3.90 and 7.69); *Daniels-Feasel v. Forest Pharmaceuticals, Inc.*, No. 17-cv-4188, 2021 WL 4037820, at *8 (S.D.N.Y. Sept. 3, 2021), *aff'd* No. 22-146, 2023 WL 4837521 (2d Cir. 2023) (risk ratio of 2.2). The district court next found that Baccarelli improperly found support for the strength criterion in meta-studies that synthesized underlying studies with varying levels of statistical significance. For example, the district court noted that while the Alemany 2021 meta-study found an overall risk ratio of 1.19 for ASD, only one of the six underlying studies it considered found a statistically significant risk ratio. The court concluded from this that it was "misleading to characterize the highly heterogenous body of literature as reporting consistent statistically significant associations." 707 F. Supp. 3d at 347.

But Baccarelli explained that while the risk ratio reported by each study is relatively low, ranging from 1.0 to 2.0, there is "no general rule for how large an association needs to be" to satisfy this criterion. App'x 1901; *see also Hardeman v. Monsanto Co.*, 997 F.3d 941, 966 (9th Cir. 2021) (rejecting contention that plaintiff's experts were wrongfully admitted because they failed "to present a study with an adjusted odds ratio above 2.0" and recognizing that "a hardline increase in a risk

[20] A risk ratio refers to "[t]he ratio of the risk of disease or death among people exposed to an agent to the risk among the unexposed." RMSE at 1019. For example, a risk ratio of 2.0 means that "the disease occurs twice as frequently among" people exposed to the agent as compared to those not exposed. *Id.*

statistic, or even an adjusted odds ratio above 2.0, is [not] necessary for finding a strong association"), *abrogated on other grounds by Monsanto Co. v. Durnell*, No. 24-1068, 2026 WL 1825691 (June 25, 2026). He further noted that the risk ratio for ADHD and ASD is stronger than others that epidemiologists have generally agreed plausibly reflect causal effects. Among these are the links between air pollution and mortality, between smoking and heart disease, and between secondhand smoke and lung cancer. App'x 1901. In other words, Baccarelli demonstrated that it is not inconsistent with epidemiological practice to conclude that the strength condition is satisfied with a risk ratio between 1.0 and 2.0. Moreover, Baccarelli correctly stated that it is not inconsistent with epidemiological practice to cite a meta-study—the express purpose of which is to "provide a more precise estimate of the true effect size of an association, by synthesizing data from multiple studies and reducing random error"—in support of a causal conclusion, even if the individual cohorts in that meta-analysis do not produce statistically significant results. App'x 1807.

Baccarelli additionally testified that the strength criterion is further supported because the magnitude of risk measured in many of the studies he assessed may have been dampened due to their reliance on maternal self-reporting of acetaminophen, which likely underestimated the true exposure. App'x 1901. In support of this conclusion, he cites Baker 2020[21], a study that directly measured the presence of acetaminophen in infant meconium (an infant's first feces that can

[21] Baker et al., *Association of Prenatal Acetaminophen Exposure Measured in Meconium with Risk of Attention-Deficit/Hyperactivity Disorder Mediated by Frontoparietal Network Brain Connectivity*, 174(11) JAMA Pediatrics 1073 (2020).

reflect prenatal exposure) and which reported a risk ratio of 2.43. The district court criticized Baccarelli for relying on Baker 2020, despite recognizing it as "well-regarded," because that study had a "wide confidence interval" (spanning 1.41 to 4.21), which the district court held "undercut[] its reliability." 707 F. Supp. 3d at 348. But although the width of a confidence interval is relevant to the reliability of a risk ratio estimate, that does not warrant the wholesale rejection of Baker 2020. Rather, that is a fact for a factfinder to consider in assessing the persuasiveness of the study.

Finally, on specificity, the district said that in light of the "complexity" suggested by the "absence" of specificity, Baccarelli neglected to "consider whether the studies upon which [he] is relying adequately considered confounding effects, among other things." 707 F. Supp. 3d at 349. That gives more weight to the specificity factor than many epidemiologists do. Baccarelli explained that even if the specificity criterion is not satisfied, it is considered "all but irrelevant" to modern epidemiologists. App'x 1904; *see also* RMSE at 985 (describing the specificity factor as "highly debatable"); VanderWeel et al., *Causal Inference and Scientific Reasoning*, in Lash's Modern Epidemiology 67 (4th ed. 2020) (recognizing that "many behavioral, environmental, social, and genetic risk factors have been causally linked to more than one health outcome"). The district court cited no authority for the proposition that, in a Bradford Hill analysis, lack of specificity necessarily indicates a dubious causal analysis. Where, as here, the causes of a disorder are multifactorial, lack of specificity does not weigh significantly against an ultimate conclusion of plausible causality. *See* App'x 1904.

ii. Conclusions on contested scientific questions

The second group of errors in the district court's analysis of Baccarelli's Bradford Hill testimony concerns instances in which the district court arrives at a definitive interpretation of a given study that reasonable scientists could interpret, and indeed have interpreted, differently. The district court's analysis of these issues is couched in terms of scientific "rules" of interpretation that it decided Baccarelli did not follow: first, that he engaged in cherry-picking, and second, that he exceeded the limitations that authors placed upon their own studies. While these court-created rules address legitimate concerns pertaining to the reliability of expert testimony, the court applied them in a manner inconsistent with how scientists practice.

Take cherry-picking. Cherry-picking occurs when an expert points to data that support a particular position while ignoring or discrediting data that contradict it. It is *not* cherry-picking for an expert to prefer one study to another when he offers a coherent, scientifically plausible reason for the preference. This issue is illustrated by the district court's discussion of Baccarelli's testimony on genetic confounding (that is, whether any observed association between prenatal acetaminophen and ASD and ADHD can be attributed to genetics). No one, including Baccarelli, denies that genetics play a role in ADHD/ASD. But that does not mean that the scientific literature has concluded that genetic confounding accounts for *all* of the observed association. On this issue, reasonable scientists can, and do, disagree. The district court characterized Baccarelli's reasonable interpretations of studies addressing genetic confounding as cherry-picking. In

doing so, the district court ignored the detailed explanations Baccarelli provided for his interpretations. That was error.

The district court first took issue with Baccarelli's treatment of Brandlistuen 2013[22] and Gustavson 2021, two sibling-control studies designed to control for the effects of genetic confounding. The two studies examined the same population several years apart. Brandlistuen 2013, the earlier study, measured ADHD symptoms as its endpoint, while Gustavson 2021, the later study, measured ADHD diagnoses. Brandlistuen 2013 found an association that persisted even after controlling for genetic confounding using a method called a "sibling-control" analysis, while Gustavson 2021 did not. The district court criticized Baccarelli for interpreting these studies as supporting his conclusion that genetic confounding does not account for the entire association at issue, stating that he did not "explain his disparate treatment of the two studies." 707 F. Supp. 3d at 353. But Baccarelli did: In his report, he explains that results from sibling-control studies are biased toward the null—that is, they understate the true associations at issue—because they control for both the genetic factors that directly cause the disease outcome at issue, *and* the genetic factors by which the agent at issue may cause the disease outcome. Baccarelli also noted that the authors of Gustavson 2021 acknowledged this as a limitation in their analysis, and that this limitation "may lead to underestimation of association estimates." App'x 1860 (internal quotation marks omitted). Moreover, Baccarelli does not overstate the value of Brandlistuen 2013

[22] Brandlistuen et al., *Prenatal Paracetamol Exposure and Child Neurodevelopment: A Sibling-Controlled Cohort Study*, 42(6) Int'l J. Epidemiology 1702 (2013).

and Gustavson 2021 to his ultimate conclusion, saying only that "[i]t is unclear whether these studies provide any meaningful evidence regarding the role of unmeasured or genetic confounding." App'x 1866. It was thus error to charge Baccarelli with cherry-picking in this part of his testimony.

Another example of a misplaced charge of cherry-picking is the district court's treatment of Baccarelli's testimony concerning Ystrom 2017[23]. That study used paternal acetaminophen use as a negative control meant to isolate the effect of acetaminophen on ADHD from familial effects, since, as the Ystrom authors hypothesized, "[i]f the association is due to unobserved familial factors (e.g., genetic factors), paternal use of acetaminophen may also be associated with ADHD in a way similar to maternal use of acetaminophen."[24] The authors found, among other results, an association between paternal preconceptual use of acetaminophen and ADHD. The district court suggested that Baccarelli's discussion of this study is not "scientifically sound" because he "speculates that paternal use might have [instead] been serving as an imperfect proxy for maternal use because, among other things, 'fathers and mothers often share medications and medicine cabinets.'" 707 F. Supp. 3d at 353, quoting App'x 1858. That speculation, the district court decided, was another instance of Baccarelli's "dismissal of evidence that challenges his thesis." *Id.*

---

[23] Ystrom et al., *Prenatal Exposure to Acetaminophen and Risk of ADHD*, 140(5) Pediatrics 1 (2017).

[24] *See, e.g.*, Sanderson et al., *Negative Control Exposure Studies in the Presence of Measurement Error: Implications for Attempted Effect Estimate Calibration*, 47(2) Int. J. Epidemiology 587 (2018) (discussing negative control experiments).

But the district court's charge misrepresents the nuance with which Baccarelli addressed this study. Baccarelli explained, and the authors of Ystrom 2017 acknowledged, that "the mechanics of the ADHD effect of paternal acetaminophen use before pregnancy are unclear but . . . may be due to male germ-line epigenetic effects as described in endocrine disruption effects of acetaminophen on the human testis." App'x 1858 (internal citations omitted). In other words, paternal preconceptual acetaminophen use may itself also be causally associated with ADHD in a manner that suggests that it should not operate as a valid negative control. And although Baccarelli's reference to shared medicine cabinets may not have used the most "scientific" language, his underlying point is not unreasonable. The study design used paternal acetaminophen exposure as a negative control but also found that paternal use was correlated with maternal use—so, as Baccarelli explains, if paternal use is actually acting as a proxy for maternal use, then the association that was found between paternal use and ADHD could be consistent with a causal relationship between maternal use and ADHD. Baccarelli's consideration of this study is not an example of results-driven analysis of the sort that would render his testimony unreliable. Again, we emphasize that we are not saying that the district court's questioning of Baccarelli's logic is itself unreasonable; still less are we suggesting that similar arguments that may be made by Defendants-Appellees' experts (which are not before us for review, and the admissibility of which has not been decided by the district court) would be improper, unreliable, or "unscientific." The point is that these are arguments and counter-arguments of the sort that scientists can and do make when engaging in professional academic debates about contested

scientific questions. The gatekeeping role of the district court is to shield the jury from testimony by even qualified experts that is outside the boundaries of ordinary scientific discourse and thus constitutes "junk science." It is not to adjudicate which points in the debate are more persuasive; that is the role of the jury.[25]

The second way that the district court censures Baccarelli for drawing reasonable conclusions is by misapplying the rule that when an expert "relies on the studies of others, he must not exceed the limitations the authors themselves place on the study." *Daniels-Feasel*, 2021 WL 4037820, at *4 (internal quotation marks omitted). It is important that when experts use work produced by other scientists, they do not paper over the limitations inherent in those studies. But the district court stretches that maxim too far in the context of a Bradford Hill analysis, which is a methodology for synthesizing many scientific results where no single one definitively establishes or negates a causal relationship. A study that does not itself prove causation, because of its own limitations, can nonetheless serve as a reliable link in a chain supporting a particular Bradford Hill factor. And that study,

[25] This understanding is fully consistent with the 2023 amendments to Rule 702. Those amendments sought to correct the "hands-off" approach to expert testimony that the Rules Committee perceived had been applied by many district courts. *See* Fed. R. Civ. P. 702 advisory committee's note to 2023 amendment (noting that many district courts improperly treated "critical questions of the sufficiency of an expert's basis, and the application of the expert's methodology" as "questions of weight and not admissibility"). But the principle that a district court must take a hard look at the experts' applications of accepted methods does not justify disregarding another principle—that the district court cannot decide a *Daubert* motion based on its own judgment of an expert's persuasiveness. Instead, the district court should strike a middle-ground between those two principles, taking a hard look at the expert's opinions to ensure that they are reliable without necessarily addressing whether they are correct. We recognize that this is easier said than done, but we trust the district courts to perform this inquiry and afford them substantial discretion in doing so.

in combination with other studies that address that limitation (and indeed may have other deficiencies of their own), may ultimately support a conclusion that points toward causation. This issue is best illustrated by the district court's treatment of the consistency and strength Bradford Hill factors allegedly supporting Baccarelli's conclusion.

On the issue of consistency, the district court concluded that limitations in each of the studies on prenatal acetaminophen use and ASD precluded Baccarelli from finding that the consistency criterion is satisfied for ASD. 707 F. Supp. 3d at 343–44. In doing so, however, the district court unreasonably discounted Baccarelli's explanation of why those limitations should not bar a finding of consistency. And it ignored the opinions of the authors of those studies, and of independent scientists who reviewed those studies, that they could support a finding of consistency.

Take the first of these studies, Ji 2020, which measured the association between acetaminophen detected in the umbilical cord at birth and ASD diagnosis and found a statistically significant increase in ASD diagnoses among children in the highest tertile of acetaminophen exposure, compared to the lowest. *Id.* at 343. Baccarelli did acknowledge the limitations of this study: for example, that acetaminophen detected in the umbilical cord at the time of birth would reflect only peripartum exposure (*i.e.*, exposure at the time of or in the hours before birth), and acetaminophen was detected in 100% of mothers measured, "which contradicts the common knowledge that only about half of pregnant women take acetaminophen, and certainly much less during the peripartum period." App'x

1833. That, according to Baccarelli, could mean that that the study included false-positives based on environmental exposures to acetaminophen (e.g., in drinking water) or that the authors included values detected below the assay's detection limit (i.e., the minimum concentration that can, in a statistically significant manner, be distinguished from zero). *See* App'x 1843–44. But he went on to point out why the study's limitations would not detract from interpreting the study as some support for the consistency criterion. For example, he explained that the measurement of acetaminophen "may not fully reflect prenatal exposure to acetaminophen"—which would mean that Ji 2020 might actually *underestimate* the effect of total prenatal exposure. App'x 1844. And on the issue of extra-sensitive acetaminophen detection, he noted that because the authors performed a tertile analysis on low, medium, and high exposure, false positives with minimal levels of exposure would have been grouped together in the low category—which would have protected against bias introduced by those false positives. *See* App'x 1844.

To be clear, the limitations in Ji 2020 are not insignificant and certainly go to the weight that the fact-finder may attribute to Baccarelli's testimony on this evidence. But they do not render the study "largely irrelevant," as the district court suggests. 707 F. Supp. 3d at 343. Nor do they prevent Baccarelli from interpreting the study as supportive of the consistency criterion. Baccarelli's reliance on Ji 2020, like the rest of his opinion, can be tested by cross-examination, contested by opposing experts who may evaluate the evidence differently, and forcefully advocated for and against by the parties' attorneys. We hold only that Baccarelli's report acknowledges the limitations of the study and gives a scientific explanation

of his reasons for finding the study consistent as demanded by the accepted Bradford Hill methodology.

The second study, Liew 2016a[26], found an association between acetaminophen and ASD with hyperkinetic disorder ("HKD"), a condition recognized under the World Health Organization's International Classification of Diseases that is analogous to a severe ADHD diagnosis under the DSM. *See* Liew 2016a at 952. That study too has limitations. For example, it found no association between acetaminophen and ASD without HKD. And in its analysis of trimester-specific exposure, it found statistically significant associations for acetaminophen use in all three trimesters, and the first and second trimesters combined, but not the second and third combined or first and third combined. Such limitations, the district court reasoned, constitute conflicting results that undermine consistency. *See* 707 F. Supp. 3d at 343–44. But again, Baccarelli addressed those limitations. And for reasons he explains, he concluded that they did not preclude a finding of consistency with other literature. For example, he adopted the authors' own acknowledgment that ASD with hyperkinetic features may be a subtype of ASD, rather than a different diagnosis altogether. App'x 1841. And, with respect to the issue of variable trimester-specific results, Baccarelli explained that "a set of results is consistent even if some of the results are not statistically significant." App'x

[26] Liew et al., *Maternal Use of Acetaminophen During Pregnancy and Risk of Autism Spectrum Disorders in Childhood: A Danish National Birth Cohort Study*, 9 Autism Rsch. 951 (2016). The district court referenced three articles published by Liew in 2016. We need only address the one it referred to as "Liew 2016a."

1769.[27] Indeed, Liew 2016a found uniformly positive associations with acetaminophen exposure by trimester and ASD, ranging from 1.03 (first trimester exposure only) to 1.39 (exposure in all three trimesters). The varying statistical significance of those results does not necessarily render them inconsistent for the purpose of the Bradford Hill analysis. As a result, it was not improper for Baccarelli to cite Liew 2016a in support of the consistency criterion.

In the end, Baccarelli's testimony on consistency is consistent with the findings of the studies that he cites. For example, the authors of Ji 2020 noted that its findings "support previous studies regarding the association between prenatal and perinatal acetaminophen exposure and childhood neurodevelopment risk." Ji 2020 at 180. The authors of Avella-Garcia 2016, another study he examines, conclude that their findings "agree with reports of an association between prenatal exposure to acetaminophen and ADHD behaviours, diagnosis or medication use in childhood." Avella-Garcia 2016 at 1993. And the authors of Liew 2016a noted that their findings built upon earlier work showing that "prenatal exposure to acetaminophen may increase[] the risk of developing hyperkinetic disorders and social and behavioral problems." Liew 2016a at 954. That the above scientists operating outside the courtroom also interpret these studies as consistent with other findings of an association supports Baccarelli's conclusion that the consistency criterion is sufficiently satisfied.

---

[27] *See also* App'x 1799–1800 ("[E]pidemiologists are often less concerned with the precise p-value [which may suggest a yes/no answer that cannot be inferred from a single study] and more interested in understanding whether the results are consistent with what we know about the biology of the disease or exposure being studied.").

To be clear, an expert's testimony is not reliable simply because other scientists have expressed the same view in a peer-reviewed publication. Just as evidence of disagreement among scientists as to the correct interpretation of a body of evidence does not necessarily render an expert's opinion unreliable, expressions of support for an opinion, without more, do not necessarily qualify the opinion as reliable. Rather, the district court must take a "hard look" at the literature an expert cites to determine whether that literature actually reflects how experts operate in the field or if it merely presents a one-off opinion that the field has roundly rejected or would be likely to. And that, we believe, the district court did not adequately do by substituting its own judgment of the evidence on an issue where scientists may reasonably disagree.

Further, the district court similarly censured Baccarelli's discussion of the strength factor. There, Baccarelli opined that because the magnitude of the risk ratio in many studies may have been dampened due to reliance on maternal self-reporting, the strong association found in Baker 2020—which directly measured acetaminophen levels in the child's meconium—provided strong evidence in support of the strength criterion. The district court noted, however, that that study "had not controlled for confounding by either indication or genetics." 707 F. Supp. 3d at 348. And it held that Baccarelli's testimony on the strength criterion was insufficient in part because he failed to examine that limitation.

There are two problems with that holding. First, Baccarelli *did* address the issue of confounding in the Baker 2020 study. He explained that the effect size found in Baker 2020 is "so large" that "residual confounding by uncontrolled

factors is a less likely explanation for the identified associations"—a conclusion that the authors of the Consensus Statement agreed with. App'x 1902 (internal quotation marks omitted). And second, while Baccarelli did not include an extensive discussion of genetic confounding in his analysis of the strength criterion, he did perform that analysis elsewhere in his expert testimony. For that reason, it does not surpass the limitations in Baker 2020 to conclude that it, in combination with other studies that do control for confounding, can be relied on in support of a plausible causal relationship.

For the above reasons, we conclude that the district court erred in excluding Baccarelli's testimony. While the district court's reasoning was considered and extensive, its analysis frequently overstepped its gatekeeping function by substituting its own judgments about the requirements of causality and the persuasiveness of various studies for those of scientists operating in the field. As a result, it ignored that Baccarelli acted in accordance with how scientists approach the question of causation using the Bradford Hill method and have previously interpreted the literature upon which he relied. When an expert reliably applies a scientifically accepted methodology on an issue that is reasonably subject to debate, it is for the jury to decide the opinion's correctness.[28]

[28] Because we conclude that Baccarelli's Bradford Hill analysis is admissible, we need not address Baccarelli's Navigation Guide, the application of which yielded the same conclusion. We express no opinion on the reliability of the Navigation Guide as a methodology or Baccarelli's application of it.

2. Hollander

Hollander's initial report focused primarily on the appropriateness of analyzing ASD, ADHD, and related NDDs together in a transdiagnostic analysis, providing support for Baccarelli's Bradford Hill analysis. In doing so, he explained the "considerable genetic, neuropsychological and clinical overlap" between the two disorders. App'x 2533. The district court excluded that opinion in its entirety.[29] It acknowledged that Hollander offered a "reliable assessment of the literature" on the overlap between ASD and ADHD but faulted him for failing to explain how a transdiagnostic Bradford Hill analysis should be structured and concluded that the studies he cited did not adequately support the conclusion that the scientific community would accept a Bradford Hill analysis that assessed ADHD and ASD together. 707 F. Supp. 3d at 364. But, as explained above, scientists have *in fact* analyzed the disorders together, and Hollander's discussion, while not necessary to justify Baccarelli's transdiagnostic approach, provides helpful context as to the relationship between ASD and ADHD.

In addition, the district court dismissed Hollander's reliance on certain scholarship explaining the value of a transdiagnostic perspective because those articles were "irrelevant" to the "transdiagnostic Bradford Hill analyses presented by" Baccarelli and Cabrera. 707 F. Supp. 3d at 366. But the district court adopted an unduly narrow definition of relevance. For example, Hollander cites

[29] Hollander also offered a biological plausibility opinion in his opening report and a Bradford Hill analysis in his rebuttal report, both of which the district court excluded. Plaintiffs-Appellants do not appeal those exclusions here, so we do not discuss them.

Vandewouw 2023[30] in support of considering ADHD and ASD together in a single causal analysis. Proceeding from the understanding that ASD, ADHD and OCD "are highly heterogenous in biology and phenotype [i.e., observable characteristics] within each condition and significantly overlapping," the study reviewed brain scans of individuals diagnosed with those conditions and found that differences in brain function corresponded to differences in behavior (e.g., hyperactivity and impulsivity) but not differences in clinical diagnoses. Thus, the authors concluded that their finding put pressure on the field's reliance "on diagnostic categories, which have increasingly been shown not to reflect distinct biological and phenotypic constructs." The district court criticized Hollander's reliance on the study because, while it is generally supportive of his position, it did not establish in definitive terms that proof of a causal relationship between acetaminophen and ADHD suffices as proof of a causal relationship between acetaminophen and ASD. 707 F. Supp. 3d at 366. But as Hollander's testimony shows, scientists in this field require no such proof to conduct a transdiagnostic analysis. As he explains, the transdiagnostic approach may be appropriate when, as here, "[t]he biological factors behind the[] symptoms cut across traditional diagnostic boundaries" such that "a holistic and transdiagnostic approach . . . is necessary to fully understand the highly heterogenous conditions" at issue. App'x 2482–83. The district court therefore overstepped its gatekeeping function in

[30] Vandewouw et al., *Identifying Replicable Subgroups in Neurodevelopmental Conditions Using Resting-State Functional Magnetic Resonance Imaging Data*, 6(3) JAMA Network Open: e232066 (2023).

excluding Hollander's testimony on the appropriateness of considering ASD, ADHD, and NDDs in a single Bradford Hill analysis.

3. Pearson

Pearson opined that a causal relationship between prenatal acetaminophen exposure and ADHD and ASD is consistent with existing biological knowledge based on preclinical literature, focusing primarily on animal studies. He concluded that the majority of preclinical studies demonstrate that acetaminophen has disruptive effects on neurodevelopment through multiple independent and interacting mechanisms, and that these studies support findings of a possible causal association between prenatal acetaminophen use and ADHD and ASD observed in epidemiological studies of humans. In his report, he acknowledges inconsistencies among the studies that he analyzes, including in the directionality of the effects observed—for example, that acetaminophen exposure may cause an increase in a particular behavior among some subjects, and a decrease in that behavior among other subjects. He explains, though, that "the heterogeneity of the results is not a reason to dismiss the effects of APAP shown in these individual studies," because "neurodevelopmental perturbation of prenatal APAP can manifest in various ways in terms of directionality." App'x 2144. The district court decided that this conclusion rendered Pearson's testimony unreliable because it amounted to an unreasonable conclusion that "heterogeneity and outright inconsistency of results don't ultimately matter," and because his conclusion to this effect was not adequately supported. 707 F. Supp. 3d at 369.

The district court erred in excluding Pearson's testimony on these grounds. Admittedly, it seems counterintuitive, for example, that a study observing a *decrease* in hyperactivity among acetaminophen-dosed rodents may evidence that the same treatment causes an *increase* in ADHD among humans. But Pearson's point is not unreasonable: Any behavioral evidence of neurodevelopmental change attributable to acetaminophen supports the conclusion that acetaminophen disrupts neurodevelopment. And his testimony to that effect is more nuanced and limited than the district court concludes.

First, Pearson was not without support for his contention. He cited Tyl 2008[31], a guidance document that provides a framework for reliably interpreting animal studies. That framework explains, as Pearson quotes, that "[v]ariability" in test results "can be interpreted as an endpoint itself rather than an indicator of study quality or reliability, and may be the first or most sensitive indicator of a response." App'x 2096.

Moreover, Pearson does not suggest that such inconsistent results nonetheless provide support for a causal conclusion. Rather, Pearson asserts that when such inconsistencies appear, it is appropriate "to try to understand and explain the reasons for them, possibly deciding if more than one answer to the formulated problem is plausible." App'x 2098. This is consistent with Tyl 2008's guidance, which Pearson cites, that "[p]roper interpretation involves expert

[31] Tyl et al., *Identification and Interpretation of Developmental Neurotoxicity Effects: A Report from the ILSI Research Foundation/Risk Science Institute Expert Working Group on Neurodevelopmental Endpoints*, 30 Neurotoxicology and Teratology 349 (2008).

judgment that takes into account both the biological and statistical significance of the results." App'x 2096, 4875.

Pearson took exactly that approach in evaluating the studies. For example, in his analysis of in vitro/ex utero studies examining the relationship between acetaminophen exposure and neurotoxicity, he identified three studies that found no evidence of neurotoxicity and five studies that found evidence of neurotoxicity. App'x 2140. He concluded that, overall, those studies supported an association between acetaminophen exposure and neurotoxicity. But he did not ignore the studies that seemed to contradict that conclusion. For example, he explained that one study that found no association failed to evaluate metabolized acetaminophen, which provides a separate hypothesized mechanistic pathway. App'x 2135. Another study found no neurotoxic outcomes in rat cortical cultures where APAP was administered on the third day in vitro. Pearson noted, however, that the relevant cortical cultures do not mature until at least the seventh day in vitro, such that testing on the third day is "potentially too early or at least might need to be interpreted carefully." App'x 2137. And the third study finding no neurotoxicity was designed to test the effect of other chemicals, with acetaminophen acting as a putatively non-neurotoxic control, such that the treatment dose size was too small to prompt neurotoxic effects. App'x 2137. Whether Pearson's conclusions are ultimately persuasive is not for us to decide. What we do decide is that he cannot fairly be found to have ignored evidence contrary to his conclusions, or engaged in unscientific sophistry, where he did discuss the contrary evidence and presented his reasons for discounting it, simply because a court disagrees with those reasons.

4. Cabrera

Cabrera testified on the reproductive, developmental, and neurodevelopmental toxicity of prenatal acetaminophen exposure. First, Cabrera described a proposed biological mechanism by which acetaminophen could cause ADHD and ASD. Cabrera based this mechanism on an Adverse Outcome Pathway ("AOP") addressing oxidative stress. An AOP is a model which, in Cabrera's case, was published by the Organization for Economic Co-operation and Development ("OECD"). It describes a logical sequence of causally linked events at various levels of biological organization, beginning with exposure to a particular stressor and leading to an adverse health effect.[32] Cabrera first assessed the evidence for each step in the causal pathway using a weight-of-the-evidence approach and then opined that the association between prenatal acetaminophen exposure and ADHD and ASD was biologically plausible. Plaintiffs-Appellants did not fully brief the district court's exclusion of that testimony, so we do not address it here.[33]

Next, Cabrera opined that each of the Bradford Hill criteria, with the exception of specificity, were satisfied, thereby supporting a determination of

---

[32] *See* OECD, *OECD Series on Adverse Outcome Pathways*, https://www.oecd.org/en/publications/oecd-series-on-adverse-outcome-pathways_2415170x.html. Other entities utilize AOP models as well. *See, e.g.*, EPA, *Adverse Outcome Pathways*, https://www.epa.gov/chemical-research/adverse-outcome-pathways (describing how the EPA uses AOP models); National Toxicology Program, U.S. Dep't Health & Hum. Serv., *Adverse Outcome Pathways*, https://ntp.niehs.nih.gov/whatwestudy/niceatm/comptox/ct-aop/aop.

[33] The district court determined that Cabrera's weight-of-the-evidence analysis was applied unreliably as to (1) the initiation of the AOP by acetaminophen exposure and (2) the connection between the final step in the AOP and diagnosis with ADHD and/or ASD. Plaintiffs-Appellants addressed the first gap in their appeal to this Court but not the second—so even if we were to reinstate the first link, the testimony could remain excluded.

causality. The district court determined that this testimony was unreliable because Cabrera failed to address the relationship between the factors or explain the weight that he attached to each one. The district court was within its discretion in doing so.

A scientist need not find every one of the nine Bradford Hill criteria satisfied in order to opine reliably that a causal relationship exists. For example (and as explained above with regard to Baccarelli), if found to be present, the specificity criterion may be especially powerful evidence of causality, but its absence does not exclude a finding of plausible causality. Similarly (and as also explained above), while an especially strong association between an exposure and outcome may serve as strong evidence of a causal relationship, a weak association doesn't necessarily mean that no causal relationship is present. That is particularly so when there is a good explanation for the observed magnitude—such as broad exposure to the toxin of interest, as in the case of air pollution or secondhand smoke, or mediating factors that contribute to the ultimate outcome in some but not all cases, such as genetic predispositions to a particular outcome. Because of these nuances in the interpretation of the various Bradford Hill criteria, the factors do not function as independent check-boxes, which are added to indicate the strength of a causal relationship. Instead, as multiple courts have recognized, expert analysis is necessary to understand how a particular combination of criteria does or does not support the ultimate causal conclusion. *See, e.g.*, *Daniels-Feasel*, 2023 WL 4837521, at *2, quoting *Zoloft*, 858 F.3d at 796 ("To ensure that the Bradford Hill/weight of the evidence criteria 'is truly a methodology, rather than a mere conclusion-oriented selection process . . . there must be a scientific method

of weighting that is used and explained.'"); *In re Mirena (No. II)*, 341 F. Supp. 3d at 248 ("[W]here experts have claimed to apply Bradford Hill, courts have insisted on a clear explication of the weighting assigned to the different criteria.").

That analysis is particularly necessary to enable a factfinder to "conduct[] a meaningful and informed review." *In re Mirena (No. II)*, 341 F. Supp. 3d at 249. Were a factfinder to disagree with any one of Cabrera's conclusions, for example, she "would then need to assess whether to find general causation in the absence of [that] factor[]." *Id.* By not "explain[ing] the relationship among the factors in his analysis, however," Cabrera "disable[s] such an inquiry." *Id.* at 248. As a result, the district court was permitted to conclude that Cabrera's application of the Bradford Hill method was unreliable and therefore inadmissible under Rule 702.[34]

5. Louie

Louie discussed the dose and duration at which prenatal exposure to acetaminophen increases the risk that a child will develop ADHD and/or ASD. He concluded that the epidemiological evidence provides a "clear answer": exposure "for at least 28 cumulative days during pregnancy increases the risk of ASD/ADHD development by two-fold as compared to offspring with no exposure to acetaminophen." App'x 2737, 2739. The district court excluded Louie's testimony on the grounds that he misstated certain evidence upon which he relied

[34] Plaintiffs-Appellants admittedly offer some support in the record for the proposition that scientists do not necessarily regard such weighing as mandatory, but we cannot say that the district court exceeded its discretion in relying on a proposition amply and directly supported by caselaw and that is meant to help a factfinder evaluate an expert's opinion.

and failed to justify adequately the conclusions that he drew from his review of the literature. The district court did not abuse its discretion in doing so.

First, Louie did not treat the studies he examined with requisite care. For example, he concluded that "studies by Brandlistuen et al, Ystrom et al, and Gustavson et al[] found that acetaminophen exposure beyond 28 days showed a two-fold increased risk for childhood ADHD and ASD diagnosis." App'x 2757. Not so. Ystrom and Gustavson examined the relationship between acetaminophen exposure and ADHD diagnosis, but not ASD. And Brandlistuen did not examine diagnoses of either ASD or ADHD but instead looked at the relationship between acetaminophen exposure and neurodevelopmental outcomes that are themselves connected to ASD and ADHD. As discussed earlier, it is not necessarily unreliable to consider ADHD and ASD together, or to examine symptomatic endpoints as well as diagnostic ones. But it is proper to exclude testimony that mischaracterizes a study.

Louie also failed to substantiate significant assumptions on which his testimony rested. This error is most apparent in Louie's conclusion pertaining to the relevant exposure threshold. Louie testified that Brandlistuen 2013, Ystrom 2017, and Gustavson 2021 "all found that acetaminophen exposure beyond 28 days showed a two-fold increased risk for childhood ADHD and ASD diagnosis." App'x 2757. But that is either a misinterpretation of those studies' findings, or an impermissibly imprecise extension of them. Each of those studies examined the frequency of the measured outcome among mothers who ingested acetaminophen for more than 28 or 29 days, compared to those who did not ingest any

acetaminophen. The "twofold risk" that each study found, therefore, was not the risk associated with ingesting acetaminophen with *any* frequency beyond the twenty-eight day threshold, as Louie concludes. It was, instead, the average risk associated with prenatal acetaminophen exposure among the group of mothers who ingested acetaminophen for between twenty-eight and two-hundred eighty days, the length of an entire pregnancy. As a result, the district court was entitled to exclude Louie's testimony because "there is simply too great an analytical gap between the data and the opinion proffered." *Joiner*, 522 U.S. 136, 146 (1997).

***

As explained above, the district court erred in dismissing expert testimony from Baccarelli, Hollander, and Pearson. It therefore follows that the district court also erred in granting summary judgment for Defendants-Appellees on the basis that Plaintiffs-Appellants lacked any evidence of general causation.

C. Preemption

As an alternative basis for affirmance, Defendants-Appellees argue that federal law preempts the failure-to-warn claims raised here. For the reasons articulated by the district court in its excellent analysis of the issue, we do not agree.

The Supremacy Clause provides that federal law "shall be the supreme Law of the Land." U.S. const. art. VI, cl. 2. Consistent with that command, "[w]here federal and state law conflict—that is, where it is impossible for a party to follow both federal and state law—state law must give way." *Gibbons v. Bristol-Myers Squibb Co.*, 919 F.3d 699, 708 (2d Cir. 2019). "When addressing federal preemption

questions, 'we have long presumed that Congress does not cavalierly pre-empt state-law causes of action.'" *Marentette v. Abbott Laboratories*, 886 F.3d 112, 117 (2d Cir. 2018), quoting *Medtronic, Inc. v. Lohr*, 518 U.S. 470, 485 (1996). We therefore "start with the assumption that the historic police powers of the States [are] not to be superseded . . . unless that [is] the clear and manifest purpose of Congress." *Wyeth*, 555 U.S. at 565 (internal quotation marks omitted).

Defendants-Appellees contend that the FDA's Pregnancy Warning Regulation creates an exclusive process for adding pregnancy-related warnings to drug labels. Thus they assert that they could give a specific warning on ADHD and ASD if, and only if, the FDA requires one through the IAAA monograph or an NDA. But this is plainly not what the text of the Pregnancy Warning Regulation commands. The Regulation requires manufacturers to display the general warning or, when one is provided by the FDA, a specific warning. It does not prohibit manufacturers from adding *supplemental* warnings when additional, pregnancy-related risks become known. And it does not modify the manufacturer's ultimate responsibility for the accuracy and adequacy of their label.

Defendants-Appellees look next to the Exact Language Regulation for support. But the Exact Language Regulation applies only where exact language has been specified, and as such, has no bearing on the addition of supplemental warnings. Defendants-Appellees can readily comply with the Pregnancy Warning Regulation and Exact Language Regulation, while warding off any failure-to-warn claims, by providing the general pregnancy warning verbatim, as specified in 21 C.F.R. § 201.63, and a supplemental warning about any additional risk of ADHD

and ASD. "Impossibility pre-emption is a demanding defense," *Wyeth*, 555 U.S. at 573, and Defendants-Appellees have not shown that simultaneous compliance with the federal Pregnancy Warning Regulation and state law warning requirements is impossible.[35]

D. Plaintiffs-Appellants' Remaining Arguments

Plaintiffs-Appellants in *Phippen* make two additional arguments. First, they contest the exclusion of their expert, Ness, whom they introduced to testify to general causation and ADHD after the district court excluded Baccarelli, Hollander, Pearson, Cabrera, and Louie. Second, they argue that even if all their

[35] Defendants further argue that the history of the Pregnancy Warning Regulation shows that it preempts any requirements under state law. When the FDA first introduced the regulation in 1982, it explained that competing regulations from other states could "weaken [its] efforts to develop comprehensive national labeling and other requirements for OTC drugs" and wanted to ensure that manufacturers could "use the new FDA labeling without also being required to use the pregnancy-nursing warning required by any State." *Pregnant or Nursing Women*, 47 Fed. Reg. 54750, 54756–57 (Dec. 3, 1982) Admittedly, those statements indicate that the agency may have thought it wise for the Pregnancy Warning Regulation to preempt other state-law requirements, but the plain language certainly does not require that conclusion. *See Safe Haven Home Care, Inc. v. United States Dep't of Health & Hum. Servs.*, 130 F.4th 305, 315 (2d Cir. 2025) (explaining that "we need not defer to the agency's interpretation of its own regulation" when "the regulation is unambiguous"). Moreover, the agency's subsequent actions appear to disavow any such intention. As noted above, in 1999, the agency issued a comprehensive rule governing OTC drug labeling, which, in part, amended the Pregnancy Warning to its present form. *See Over the Counter Drugs*, 64 Fed. Reg. at 13286. Prior to issuing the final rule, the agency had considered including a provision that would "preempt State and local rules that establish different requirements than those in the proposed rule, to promote a national, standardized format for all OTC drug product labeling." *Id.* at 13254. It ultimately decided not to enact such a rule because Congress, in the interim, had passed the Food and Drug Administration Modernization Act of 1997 ("FDAMA"). That new law included a provision governing preemption, which the FDA decided it would rely on "in addressing preemption issues." *Id.* at 13272, citing 21 U.S.C. § 379r. The provision bans states from enacting any labeling requirement "that is different from or in addition to" those required by the FDA, *but* it contains a carveout for state products liability law: "Nothing in this section shall be construed to modify or otherwise affect any action or the liability of any person under the product liability law of any State." 21 U.S.C. § 379r(a), (e). The FDA's decision to adopt the preemption approach outlined in the FDAMA thus undercuts any continued relevance of its 1982 statement.

experts remain excluded, statements made by Defendants-Appellees' expert, Faraone, which are supportive of a possible causal relationship between prenatal acetaminophen exposure and ADHD, would be sufficient to permit a jury to conclude in Plaintiffs-Appellants' favor on the issue of general causation.

Their first argument is one that, in the interest of judicial economy, we do not reach today. The *Phippen* Plaintiffs-Appellants offered Ness only after the five *Rutledge* experts were excluded, and only to support the limited proposition of a causal relationship between prenatal acetaminophen exposure and ADHD. Because Baccarelli, Hollander, and Pearson's opinions are admitted, the *Phippen* Plaintiffs-Appellants may well wish to proceed at this time without Ness's testimony. Moreover, without expressing any opinion about the correctness of the district court's rejection of Ness's testimony, to the extent the *Phippen* Plaintiffs-Appellants *do* decide to press ahead with Ness as an additional expert witness, the district court, informed by our analysis above, may well choose to revisit its conclusions about her testimony. We therefore express no opinion on its admissibility, vacate the district court's judgment in *Phippen*, and remand to allow further proceedings consistent with this opinion.

Second, the *Phippen* Plaintiffs-Appellants contend that a jury could find general causation based on certain statements made by Faraone, namely fragments from his deposition, his scientific papers, and his popular science writings that, read in isolation, could be construed as evidence in support of a causal relationship between prenatal acetaminophen exposure and ADHD. Even if all of the statements identified by Plaintiffs-Appellants are deemed admissible, this

"concatenation of scientific propositions" falls far short of reliable expert testimony upon which a jury may find general causation. *In re Mirena IUS Levonorgestrel-Related Prods. Liab. Litig. (No. II)*, 387 F. Supp. 3d 323, 354 (S.D.N.Y. June 11, 2019). The district court correctly determined that, "[e]ven if cobbled together, the Faraone [statements] do[] not constitute a Bradford Hill analysis in support of plaintiffs' thesis." *In re Acetaminophen*, 2024 WL 3874183, at *4.

III. Conclusion

In this opinion, we have assessed the admissibility, and only the admissibility, of the opinions of Plaintiffs-Appellants' experts. We have not assessed the weight they should be given by a factfinder. Defendants-Appellees' experts are not before us. We do not prejudge whether the expert testimony discussed here, when considered in light of the *totality* of the evidence, including any admissible testimony from Defendants-Appellees' experts, suffices to raise an issue of fact for summary judgment purposes. Moreover, we have considered the record as it was before the district court. Since the district court's decision, numerous relevant studies on this issue have been published. The district court may consider it prudent to invite the parties and their experts to address those new studies.

For the reasons explained above, we VACATE the district court's judgments and REMAND for further proceedings consistent with this opinion.