IN THE UNITED STATES DISTRICT COURT FOR THE SOUTHERN DISTRICT OF WEST VIRGINIA

CHARLESTON DIVISION

OHIO VALLEY ENVIRONMENTAL COALITION, WEST VIRGINIA HIGHLANDS CONSERVANCY and SIERRA CLUB,

Plaintiffs,

v.

CIVIL ACTION NO. 2:13-21588
(Consolidated with 2:13-16044)

FOLA COAL COMPANY, LLC,

Defendant.

MEMORANDUM OPINION AND ORDER

This suit concerns allegations that Defendant Fola Coal Company, LLC, has violated the narrative water quality standards of three separate permits for discharges from three mines into tributaries of Leatherwood Creek. On June 1–4, 2015, the Court held a bench trial regarding jurisdiction and liability, and the parties timely conducted post-trial briefing.

As explained below, the Court **FINDS** that Plaintiffs have established, by a preponderance of the evidence, that Defendant has committed at least one violation of its permits governing Fola Mine No. 2 and Fola Mine No. 6 by discharging into Road Fork and Cogar Hollow high levels of ionic pollution, which have caused or materially contributed to a significant adverse impact to the chemical and biological components of the applicable streams' aquatic ecosystem, in violation of the narrative water quality standards that are incorporated into those permits. However, the Court further **FINDS** that Plaintiffs have not met their burden in establishing liability for alleged

violations with respect to discharges from Fola Mine No. 4A into Right Fork, under NPDES Permit No. WV1013815.

I. BACKGROUND

Plaintiffs Ohio Valley Environmental Coalition ("OVEC"), West Virginia Highlands Conservancy, and Sierra Club filed this case pursuant to the citizen suit provisions of the Federal Water Pollution Control Act ("Clean Water Act" or "CWA"), 33 U.S.C. § 1251 et seq., and the Surface Mining Control and Reclamation Act ("SMCRA"), 30 U.S.C. § 1201 et seq. Compl., ECF No. 1. Before proceeding to the parties' evidence and arguments, the Court will first discuss the relevant regulatory framework and the factual background of this case.

A. Regulatory Framework

The primary goal of the CWA is "to restore and maintain the chemical, physical, and biological integrity of the Nation's waters." 33 U.S.C. § 1251(a). To further this goal, the Act prohibits the "discharge of any pollutant by any person" unless a statutory exception applies; the primary exception is the procurement of a National Pollutant Discharge Elimination System ("NPDES") permit. 33 U.S.C. §§ 1311(a), 1342. Under the NPDES, the U.S. Environmental Protection Agency ("EPA") or an authorized state agency can issue a permit for the discharge of any pollutant, provided that the discharge complies with the conditions of the CWA. 33 U.S.C. § 1342. A state may receive approval to administer a state-run NPDES program under the authority of 33 U.S.C. § 1342(b). West Virginia received such approval, and its NPDES program is administered through the West Virginia Department of Environmental Protection ("WVDEP"). 47 Fed. Reg. 22363-01 (May 24, 1982). All West Virginia NPDES permits incorporate by reference West Virginia Code of State Rules § 47-30-5.1.f, which states that "discharges covered by a WV/NPDES permit are to be of such quality so as not to cause violation of applicable water

quality standards promulgated by [West Virginia Code of State Rules § 47-2]."[1] This is an enforceable permit condition.[2] *See, e.g., OVEC v. Elk Run Coal Co., Inc.*, No. 3:12-cv-0785, 2014 WL 29562, at *3, *6 (S.D. W. Va. Jan. 3, 2014); *OVEC v. Elk Run Coal Co., Inc.*, No. 3:12-cv-0785, 24 F.Supp.3d 532 (S.D. W. Va. June 4, 2014); *OVEC v. Fola (Stillhouse)*, No. 2:13-5006, 2015 WL 362643 (Jan. 27, 2015).

Coal mines are also subject to regulation under the SMCRA, which prohibits any person from engaging in or carrying out surface coal mining operations without first obtaining a permit from the Office of Surface Mining Reclamation and Enforcement ("OSMRE") or an authorized state agency. 30 U.S.C. §§ 1211, 1256, 1257. A state may receive approval to administer a state-run surface mining permit program under the authority of 30 U.S.C. § 1253. In 1981, West Virginia received conditional approval of its state-run program, which is administered through the WVDEP pursuant to the West Virginia Surface Coal Mining and Reclamation Act ("WVSCMRA"). W. Va. Code §§ 22-3-1 to -33; 46 Fed. Reg. 5915-01 (Jan. 21, 1981). Regulations passed pursuant to the WVSCMRA require permittees to comply with the terms and conditions of their permits and all applicable performance standards. W. Va. Code R. § 38-2-3.33.c. One of these performance standards requires that mining discharges "shall not violate effluent limitations or cause a violation of applicable water quality standards." *Id.* § 38-2-14.5.b. Another performance standard mandates that "[a]dequate facilities shall be installed, operated and

[1] At the time of filing, the mining operations at issue here were each regulated under WV/NPDES permits reissued in 2008. At that time, W.Va. Code R. § 47-30-5.1.f read as quoted here. *See* Mem. Op. & Order 12–21, May 29, 2015, ECF No. 94.

[2] The Court pauses here to reject Defendant's recurring and spurious mischaracterizations of prior holdings. It is the express language of Defendant's federally approved WV/NPDES permits as issued by the WVDEP that requires compliance with state narrative water quality standards as an enforceable permit condition. It is the WVDEP as the issuing agency—not Congress and not this Court—that incorporated narrative water quality standards as enforceable WV/NPDES permit conditions.

maintained using the best technology currently available . . . to treat any water discharged from the permit area so that it complies with the requirements of subdivision 14.5.b of this subsection." *Id.* § 38-2-14.5.c.

B. Factual Background

This controversy concerns discharges from three surface mines along the southern portion of the Leatherwood Creek watershed: (1) Fola Surface Mine No. 2 in Clay and Nicholas Counties, West Virginia; (2) Fola Surface Mine No. 4A in Clay County, West Virginia; and (3) Fola Surface Mine No. 6 in Nicholas County, West Virginia. Stipulation ¶6, ECF No. 53.

Defendant's mining activities at Surface Mine No. 2 are regulated under WV/NDPES Permit WV1013840 and West Virginia Surface Mining Permit S201293, both originally issued in 1994. *Id.* at ¶¶6–7. WVDEP reissued WV/NPDES Permit No. WV1013840 in 2001, 2004, 2008, and 2014. At the time this complaint was filed, the 2008 reissuance was in effect. Outfall 001 of Surface Mine No. 2 discharges into Road Fork and Leatherwood Creek. *Id.*

Defendant's mining activities at Surface Mine No. 4A are regulated under WV/NPDES Permit WV1013815 and West Virginia Surface Mining Permit S200502. *Id.* at ¶¶21, 23. WV/NPDES Permit WV1013815 was originally issued in 1993, and was reissued in 1999, 2006, 2008, and 2014. At the time this complaint was filed, the 2008 reissuance was in effect. Outfalls 22, 23, and 027 of Surface Mine No. 4A discharge into Right Fork of Leatherwood Creek and Cannal Coal Hollow.[3] *Id.*

[3] Plaintiffs' Amended Complaint also complains of discharges from Outlet 025 at Surface Mine No. 4A. ECF No. 39. However, Plaintiffs did not present any testimony regarding Outlet 025 during trial. Upon Defendant's oral motion for a directed verdict regarding Outlet 025, Plaintiffs' counsel advised the Court and Defendant that they had intended to drop any claims related to Outlet 025. Tr. 3 at 204, ECF No. 106. Accordingly, any claims relating to Outlet 025 were dismissed.

Finally, Defendant's mining activities at Surface Mine No. 6 are regulated under WV/NPDES Permit WV1018001 and West Virginia Surface Mining Permit S2011999, both originally issued in 2000. *Id.* at ¶¶42–44. WVDEP reissued WV/NPDES Permit WV1018001 in 2008. *Id.* at ¶43. At the time this complaint was filed, the 2008 reissuance was in effect. Outlets 013, 015, and 017 of Surface Mine No. 6 discharge into Cogar Hollow, a small tributary of Leatherwood Creek. *Id.* at ¶43.

In recent years, water quality measurements from the above listed discharges have routinely shown discharges of high conductivity. Stipulation ¶14, ECF No. 53 (showing discharges from Outlet 001 at Mine No. 2 with conductivity measurements consistently around 3000 µS/cm); *id.* at ¶32 (showing discharges from Outlets 22, 23, 27 at Mine No. 4A consistently ranging from approximately 1500 µS/cm to above 3000 µS/cm); *id.* at ¶47 (showing discharges from Outlets 013, 015, 017 with conductivity measurements consistently ranging from approximately 2500 µS/cm to 4000 µS/cm). Water quality measurements have also revealed elevated conductivity in Leatherwood Creek and its tributaries. *Id.* at ¶13 (showing conductivity levels ranging from 3000 µS/cm to 4000 µS/cm in Road Fork); *id.* at ¶33 (showing conductivity levels consistently above 1000 µS/cm below Mine No. 4A); *id.* at ¶46 (showing conductivity levels ranging from 3000 µS/cm to 5000 µS/cm in Cogar hollow).

On June 1–4, 2015, the Court conducted a bench trial on liability issues. At the close of the evidence, the Court entered an oral finding on general causation, but reserved judgment on issues of specific causation. Tr. 4 at 259–60, June 4, 2015, ECF No. 107. Since that time, the parties have provided timely post-trial briefing. In Section II, the Court will review the evidence and arguments concerning general causation and elaborate on its general causation finding. In

Section III, the Court will move on to review the evidence and arguments concerning specific causation for each of the three mine permits at issue.[4]

[4] Defendant expresses concern regarding potential bias related to one member of the Court's staff in an early and lengthy footnote in its post-trial briefing. ECF No. 116 at 3 n.3. Specifically, Defendant provides a survey of publically available electronic media detailing the employment and advocacy history of one of the Court's term law clerks, Ms. McCrae. *Id.* Defendant refers to a denied recusal motion in a 2008 case, but does not expressly request anything from the Court on this occasion. Absent some motion, Defendant's concerns are not properly before the Court. The Court nevertheless offers the following comments.

"There is always some risk of bias; to constitute grounds for disqualification, the probability that a judge will decide a case on a basis other than the merits must be more than 'trivial.'" *United States v. DeTemple*, 162 F.3d 279, 287 (4th Cir. 1998) (citing *In the Matter of Mason*, 916 F.2d 384, 386 (7th Cir. 1990)). "A judge should not allow family, social, political, financial, or other relationships to influence judicial conduct or judgment." Canon 2, Code of Conduct of United States Judges.

In 1948, Congress authorized district courts to employ law clerks, though it did so without elaborating on the specific duties of law clerks. 28 U.S.C. § 752. Law clerks are considered part of a judge's personal staff. Code of Conduct for Judicial Employees § 310.30(a). "The proper role of the judicial adjunct, who in the federal setting may be defined as anyone who helps with the work of Article III courts but whose conditions of employment are not as prescribed in Article III, is to advise and assist the real judge. It is not to be the real judge, only called something else." *Geras v. Lafayette Display Fixtures, Inc.*, 742 F.2d 1037, 1047 (7th Cir. 1984) (Posner, J., dissenting). "No American judge today believes that a law clerk becomes a judge by preparing an opinion draft." *Id.* Judges decide cases; law clerks perform tasks as directed.

Here, Defendant appears to offer no complaint of bias on behalf of the Court, instead questioning potential biases of a subordinate staff-member, a term law clerk. Such concerns are trivial at best. At the risk of stating the obvious, a term law clerk is not a judge. A term law clerk performs tasks as delegated to him or her by a supervising judge. A term law clerk does not enjoy the exercise of discretion. That responsibility is reserved for the judge and the judge alone. A term law clerk merely acts in service of a supervising judge's discretion.

Even ignoring that obvious point, the Court further notes that whatever the specific content of individual statements, the surveyed conduct occurred over four years ago. *See DeTemple*, 162 F.3d at 287 (holding that judge in a bankruptcy fraud proceeding was not required to recuse himself, in part because the judge "last represented [defendant's creditor] almost two years before DeTemple filed for personal bankruptcy and five years prior to his indictment."). In the four years or more that has elapsed since any of the surveyed conduct, Ms. McCrae graduated with honors from a distinguished law school and completed the majority of her service to this Court. The passage of considerable time—in addition to the simple fact that a law clerk is not a federal judge—suggests that there is not a need for recusal here, a point presumably recognized by Defendant as no such motion was made.

Moreover, this Court has decided cases similar to the case at bar for nearly a decade. The decision here is consistent with existing case law. That consistent, existing case law predates not only Ms. McCrae's service as a term clerk with this Court, but also predates much of the conduct

II. General Causation

Generally speaking, Plaintiffs are faced with the dual burden of establishing both general and specific causation.[5] At the close of trial, the Court announced its finding that Plaintiffs met their burden with respect to general causation. Tr. 4 at 259–60, ECF No. 107. Specifically, Plaintiffs proved by a preponderance of the evidence that conductivity, as a measure of a consistent mix of ions typical of alkaline mine drainage in the Appalachian region, may cause or materially contribute to biological impairment to aquatic life as measured by the West Virginia Stream Condition Index ("WVSCI"), thereby constituting a violation of the narrative water quality standards incorporated into Defendant's permits.[6] *Id.*; *accord OVEC v. Elk Run Mining Co.*, 24 F.Supp. 3d 532 (S.D. W.Va. 2014); *OVEC v. Fola Coal Co. (Stillhouse)*, 2015 WL 362643 (S.D.

---

causing Defendant's concern. All told, given the nature of term clerk employment, the remoteness of the conduct complained of, and the existence of earlier, consistent precedents, it is difficult to characterize Defendant's concerns as anything more than a trivial and tangential waste of judicial resources.

[5] *See Dittrich-Bigley v. Gen-Probe, Inc.*, No. 11-1762, 2013 WL 3974107, at *7 (D. Minn. July 31, 2013) ("Generally, causation is divided into two components: general and specific. General causation is whether X *can* cause Y. Specific causation is whether X *did* cause Y.") (emphasis in original); *Heller v. Shaw Indus., Inc.*, No. 9507657, 1997 WL 535163, at *6 (E.D. Pa. Aug. 18, 1997) ("General causation addresses whether products of the same nature as [the] defendant's product are capable of causing the type of injuries alleged . . . [, while] specific causation addresses whether [the] defendant's product more likely than not caused injuries in this particular case."). *But cf. Ranes v. Adams Labs. Inc.*, 778 N.W.2d 677, 688 (Iowa 2010) (explaining that while it is analytically helpful to think in terms of general and specific causation in toxic torts cases, both are aspects of factual causation and are not necessarily separately required elements).

[6] The West Virginia Stream Condition Index, or WVSCI, is a multimetric index used to conduct biological assessments of stream conditions. The WVDEP relied on WVSCI as a means to determine whether a violation of the biological standard in subsection 3.2.i was occurring, such that a stream needed to be listed as "impaired" under a Section 303(d) of the Clean Water Act. "[T]he EPA—the final authority regarding whether a state's narrative water quality are being violated for the purposes of Section 303(d) listing—recently made the *specific finding* that WVSCI scores below 68 'indicate that [the] waters [at and in which such scores were assessed] do not achieve the West Virginia narrative criteria as applied to the aquatic life uses'—defined by the EPA to mean the biological standards embodied in § 47–2–3.23 and –3.21." *Elk Run*, 24 F.Supp.3d at 550 (quoting March 25, 2013, Letter from EPA to WVDEP).

W.Va. Jan. 27, 2015) The bases for the Court's finding on general causation are explained in further detail below, with the majority of discussion focused on Defendant's critiques of EPA's Benchmark, followed by a brief review of other scholarly publications on the question.

A. Introducing the EPA's Benchmark

Yet again, the Court begins its analysis of general causation with arguments concerning the import and reliability of the EPA's Benchmark. *See Elk Run,* 24 F.Supp.3d at 558–59; *Fola (Stillhouse),* 2015 WL 362643. In March 2011, the EPA released "A Field-Based Aquatic Life Benchmark for Conductivity in Central Appalachian Streams" ("EPA's Benchmark" or "Benchmark"). Joint Ex. 17. The EPA's Benchmark is the studied result of qualified authors and reviewers. *Id.* at ix–xiii (listing authors, contributors, and reviewers, including Defendant's expert, Dr. Charles Menzie).

In the nearly three hundred page Benchmark, the EPA reached the conclusion that "salts, as measured by conductivity, are a common cause of impairment of aquatic macroinvertebrates" in central Appalachian streams only after considering and then ruling out the potential confounding effects of habitat, organic enrichment, nutrients, deposited sediments, pH, selenium, temperature, lack of headwaters, catchment areas, settling ponds, dissolved oxygen, and metals. EPA's Benchmark at A–1, B–1; *see also id.* at A–40 ("This causal assessment presents clear evidence that the deleterious effects to benthic invertebrates are *caused by, not just associated with,* the ionic strength [, i.e., conductivity,] of the water.... When [other potential] causes are absent or removed, a relationship between conductivity and ephemeropteran [, i.e. mayfly,] richness is still evident." (emphasis added)); *id.* at A–37 ("As conductivity increases, the occurrence and capture probability decreases for many genera in West Virginia ... at the conductivity levels predicted to cause effects. The loss of these genera is a severe and clear effect."). The Benchmark also found that "of the

[nine] land uses ... analyzed, only mining especially associated with valley fills[, i.e., mountaintop mining with valley fills,] is a substantial source of the salts that are measured as conductivity." *Id.* at A–18.

The EPA ultimately concluded that the "chronic aquatic life benchmark value for conductivity" in West Virginia streams is 300 μS/cm. *Id.* at xv. To derive this recommended high-end threshold value, the EPA used the 5th percentile of a species sensitivity distribution, based on the standard methodology for deriving water-quality criteria, meaning that this 300 μS/cm benchmark value for conductivity is "expected to avoid the local extirpation [due to the salts measured as conductivity] of 95% of native species." *Id.* at xiv.

In support of both the specific 300 μS/cm benchmark value and the general causal linkage between conductivity and impairment to aquatic macroinvertebrates, the Benchmark contains a graph which charts, for 163 genera, the level of ionic exposure above which a genus is effectively absent from water bodies in a region, with conductivity readings on the x axis and proportion of genera extirpated on the y axis. *Id.* at xiv, 18 fig. 8. A fairly consistent line is formed as conductivity and extirpation both increase, illustrating the causal connection between conductivity and significant biological impairment which Plaintiffs seek to prove. *See id.* at 18 fig. 8. Relatedly, the EPA reported its finding that "the probability of impairment at 500 μS/cm is 0.72 and at 300 μS/cm, is 0.59." Tr. 2 at 111, June 2, 2015, ECF No. 100; Joint Ex. 17 at A-36. Stated differently, when conductivity reaches 300 μS/cm, it is more likely than not that the streams will suffer impairment. Moreover, the likelihood of impairment continues to increase as conductivity further exceeds that threshold. Joint Ex. 17 at A-36; Tr. 2 at 110–12, ECF No. 100.

Upon reviewing the EPA's findings, the Scientific Advisory Board ("SAB") made the follow comments:

> Mountaintop mining and valley fills are important sources of stress to aquatic systems in the Central Appalachian region, both from the perspective of localized and cumulative regional impacts. In a companion report, the Panel provides a review of EPA's assessment of the impacts associated with mountaintop mining and valley fills. There is clear evidence that valley fills are associated with increased levels of dissolved ions (measured as conductivity) in downstream waters, and that these increased levels of conductivity are associated with changes in the composition of stream biological communities.

Pls.' Ex. 128 at PE1418, Tr. 2 at 112–13, ECF No. 100. The SAB further concluded that the EPA had presented a "convincing case" for establishing the causal relationship between conductivity and loss of genera. Pls.' Ex. 128 at PE1431, Tr. 2 at 115, ECF No. 100.

Plaintiffs rely on EPA's Benchmark here as they have elsewhere: as a scientific study which, among others, supports Plaintiffs' general causation theory that high conductivity levels in streams impacted by alkaline mine drainage causes or contributes to biological impairment. Defendant's many critiques of the EPA Benchmark will be considered below. Before doing so, however, it is necessary to briefly revisit general principles regarding the degree of deference owed EPA's Benchmark in the analysis to follow.

"Particularly with environmental statutes such as the Clean Water Act, the regulatory framework ... requires sophisticated evaluation of complicated data.... [A court] therefore do[es] not sit as a scientific body in such cases, meticulously reviewing all data under a laboratory microscope." *Crutchfield v. Cnty. of Hanover, Virginia,* 325 F.3d 211, 218 (4th Cir.2003) (citation omitted) (internal quotation marks omitted). Instead, "[a] reviewing court must generally be at its most deferential when reviewing factual determinations within an agency's area of special expertise.... It is not the role of a reviewing court to second-guess the scientific judgments of the EPA." *Sw. Pennsylvania Growth Alliance v. Browner,* 121 F.3d 106, 117 (3d Cir.1997) (citation omitted) (internal quotation marks omitted); *see also Baltimore Gas & Elec. Co. v. Natural Res. Def. Council, Inc.,* 462 U.S. 87, 103, 103 S.Ct. 2246, 76 L.Ed.2d 437 (1983) ("[A] reviewing court

must remember that the [agency] is making predictions, within its area of special expertise, at the frontiers of science. When examining this kind of scientific determination, as opposed to simple findings of fact, a reviewing court must generally be at its most deferential."); *Envtl. Def. Ctr.,* 344 F.3d at 869 ("We treat EPA's decision with great deference because we are reviewing the agency's technical analysis and judgments, based on an evaluation of complex scientific data within the agency's technical expertise."); *Chem. Mfrs. Ass'n v. U.S. E.P.A.,* 919 F.2d 158, 167 (D.C.Cir.1990) ("[W]e give considerable latitude to the EPA in drawing conclusions from scientific and technological research, even where it is imperfect or preliminary." (internal quotation marks omitted)).

"[T]echnological and scientific issues ... are by their very nature difficult to resolve by traditional principles of judicial decisionmaking. For this reason, we must look at the decision not as the chemist, biologist or statistician that we are qualified neither by training nor experience to be, but as a reviewing court exercising our narrowly defined duty of holding agencies to certain minimal standards of rationality." *Reynolds Metals Co. v. U.S. E.P.A.,* 760 F.2d 549, 558–59 (4th Cir. 1985) (internal quotation marks omitted). "[A]n agency's data selection and choice of statistical methods are entitled to great deference, ... and its conclusions with respect to data and analysis need only fall within a zone of reasonableness." *Id.* at 559 (citations omitted) (internal quotation marks omitted). In the context of agency action, "if the agency fully and ably explains its course of inquiry, its analysis, and its reasoning sufficiently enough for us to discern a rational connection between its decision-making process and its ultimate decision, [a court] will let its decision stand." *Crutchfield,* 325 F.3d at 218 (brackets omitted) (internal quotation marks omitted).

In light of these precedents, and as previously analyzed by this Court, EPA's Benchmark must be afforded deference. *See Elk Run*, 24 F.Supp.3d at 558–59, *Fola (Stillhouse)*, 2015 WL 362643 at *4–6. The EPA's Benchmark methodically defines its inquiry, explains its reasonable analysis, and thoroughly supports its ultimate, rational conclusions. Additionally, the Benchmark underwent extensive scientific review, and it is respected as good—or even excellent—science within the relevant scientific community.[7] Dr. Palmer, Tr. 2 at 96, ECF No. 100.

B. Critiques of the EPA Benchmark

Turning to consider newly presented evidence and argument, two recurring questions underlie the Court's instant analysis of general causation as related to EPA's Benchmark: (1) whether specific expertise in epidemiology is required for the development or review of EPA's Benchmark; and, in a similar vein, (2) whether specific expertise in ecology is required for the development or review of EPA's Benchmark.[8] We are faced with these fundamental questions

[7] The Court further notes that sub-parts of the Benchmark were later published in an esteemed peer-reviewed scientific journal. *See* Susan M. Cormier, Glenn W. Suter II & Lei Zheng, *Derivation of a Benchmark for Freshwater Ionic Strength*, 32 Envtl. Toxicology & Chemistry 263 (2013), Joint Ex. 3; Susan M. Cormier & Glenn W. Suter II, *A Method for Assessing Causation of Field Exposure–Response Relationships*, 32 Envtl. Toxicology & Chemistry 272 (2013), Pls.' Ex. 140; Susan M. Cormier et al., *Assessing Causation of the Extirpation of Stream Macroinvertebrates by a Mixture of Ions*, 32 Envtl. Toxicology & Chemistry 277 (2013), Joint Ex. 4; Susan M. Cormier et al., *Relationship of Land Use and Elevated Ionic Strength in Appalachia Watersheds*, 32 Envtl. Toxicology & Chemistry 296 (2013), Joint Ex. 6; Susan M. Cormier & Glenn W. Suter II, *A Method for Deriving Water-Quality Benchmarks Using Field Data*, 32 Envtl. Toxicology & Chemistry 255 (2013), Pls.' Ex. 139; Glenn W. Suter II & Susan M. Cormier, *A Method for Assessing the Potential for Confounding Applied to Ionic Strength in Central Appalachian Streams*, 32 Envtl. Toxicology & Chemistry 288 (2013), Joint Ex. 5.

[8] *Compare* Dr. Garabrant, Tr. 1 at 49, ECF No. 105 (explaining his qualifications as a reviewer of the EPA Benchmark despite having no formal ecological training or experience as follows: "Well, I understand epidemiology, and I understand how to analyze data. I've spent my career doing that. I've spent my career doing divisional research, analyzing complex datasets, publishing papers in the peer-reviewed literature, and I know how to analyze data. And I do understand the principles of epidemiology. The EPA said they used epidemiology. They didn't use it properly. They did not do what scientists agree they have to do to analyze data properly. So while

because of the apparent (and unsurprising) difficultly in finding an expert in both epidemiology *and* ecology. Instead, the Court heard testimony of expert epidemiologists with no formal ecological training and expert ecologists with no formal epidemiological training.[9] From this mix of incomplete expertise, we are left with the task of sorting competing expert opinions.

As argued by Defendant, because the EPA incorporated principles of epidemiology into its causal analysis, assessing the reliability of the EPA's findings requires review by an epidemiologist.[10] Defendant's expert epidemiologist, Dr. David Garabrant, reviewed the EPA's findings and found several areas where he believed the EPA misapplied epidemiological principles. Tr. 1 at 11, June 1, 2015, ECF No. 105 (asked whether the EPA correctly applied principles of epidemiology, Dr. Garabrant responded, "In some ways, yes; and in some ways, no."). Dr. Garabrant thusly criticized perceived failures on the part of EPA (1) to consider effect modification, (2) to define reliable and valid criteria for assessing confounding, (3) to adequately respond to the quality of the available data, and (4) to transparently and non-manipulatively

---

I am not an ecologist, I do understand complex data, I do understand biostatistics, and I do understand epidemiology. The EPA is not entitled to their own version of science.") *and* Dr. Wing, Tr. 2 at 13, ECF No. 100 (explaining that "it's important to have substantive knowledge in the area in which one conducts analyses and draws opinions. . . . Because science involves more than data analysis. There's an important distinction between a data analyst and a scientist because variables in data don't speak for themselves. They require interpretation and understanding of mechanisms, theories in the field, and so on.").

[9] Meaning that whatever their qualifications, no testifying expert held a graduate-level degree in both the natural sciences *and* epidemiology, nor did any testifying expert claim expertise in both fields. That said, testifying ecological experts, e.g., Dr. Menzie, Dr. Palmer, and Dr. Baker, each have considerable expertise in statistical analysis of datasets within their respective disciplines. In contrast, Dr. Garabrant has no formal training related to aquatic ecology and testified to only reviewing literature on aquatic ecology and freshwater macroinvertebrates for purposes of this litigation. Tr. 1 at 53–55, ECF No. 105.

[10] However, also note that "epidemiological studies are not necessarily required to prove causation, as long as the methodology employed by the expert in reaching his or her conclusion is sound." *Benedi v. McNeil-P.P.C., Inc.*, 66 F.3d 1378, 1384 (4th Cir. 1995); *see also United States v. W.R. Grace*, 504 F.3d 745, 765 (9th Cir. 2007) ("the fact that a study is associational—rather than an epidemiological study intended to show causation—does not bar it from being used to inform an expert's opinion").

disclose *all* data. Each of Dr. Garabrant's critiques will be reviewed in turn below, but first the Court observes that Plaintiffs' responses to these critiques can be boiled down to the suggestion that Dr. Garabrant's critiques are fundamentally flawed insofar as Dr. Garabrant did not adequately understand the underlying subject matter, i.e., freshwater ecology.[11]

Taking a particularly illustrative example, Dr. Garabrant's critique of Table B-7 of the EPA Benchmark suggests an inability to correctly apply common statistical tools presumably well within his epidemiological expertise and, in doing so, to adequately interpret available ecological data. Table B-7 offers a relatively straightforward presentation of data purporting to represent two regression lines. Joint Ex. 17 at JE0773. Dr. Garabrant demonstrated the perceived failings of the table by attempting to recreate the graph. *See* Tr. 1 at 22–24, ECF No. 105. Using the data in the table, Dr. Garabrant's recreated graph offered only the nonsensical result of predicting a total absence of mayflies at background conductivity levels. *Id.* Basic knowledge of ecology and observed conditions tell us that, as graphed by Dr. Garabrant, the numbers in Table B-7 cannot be correct. Thus, Dr. Garabrant offered his expert opinion that the table was nonsense. Tr. 1 at 22 ("Something's seriously wrong. It is not a valid result."); *id.* at 24 ("We know that the maximum number of ephemeropteran genera is 14. Of course, the minimum has to be zero. All this [table] generates is negative numbers. It's nonsense.").

Contrary to Dr. Garabrant's opinion, Plaintiffs' expert found the same Table B-7 to be perfectly sensible with the addition of a single interpretive move. Though explained elsewhere in the Benchmark (see, e.g., Figures 13a, 13b, 13c, 13d, and 13e), the authors made no explanation of

[11] For instance, Plaintiffs' expert, Dr. Wing opined that, in his assessment, "[Dr. Garabrant's] statistical analysis and his opinions about the Benchmark are uninformed for the most part by a nuanced and thorough understanding of the topic of stream ecology, which would be necessary for a proper evaluation of that topic." Tr. 2 at 41, ECF No. 100.

whether a logarithmic scale should be used to interpret the data shown at Table B-7.[12] Though not specifying that the data in the table would need to be logarithmically transformed, according to Dr. Baker, the appropriateness of using a log scale would be obvious to an ecologist. Tr. 4 at 229–30, ECF No. 107. With that background expertise in ecological data analysis, Dr. Baker produced two graphs, each using different scales, and each showing results consistent with the data analyzed. Pls.' Exs. 176 and 177.

Thus, what appeared to one epidemiological expert to be an incorrect and nonsensical table, was in fact a perfectly sensible table that the authors merely neglected to adequately label for non-expert reviewers. While there is likely no across-the-board answer to what respective degrees of epidemiological and ecological expertise are necessary to evaluate the EPA's Benchmark, this example serves as a ready reminder in the analysis to follow that something beyond wholly non-contextual data analysis may be needed.[13]

1. Assessment of Effect Modification

[12] Dr. Baker recalled a general explanation specifying use of a logarithmic scale offered elsewhere in the Benchmark, but he was not able to quickly find such a reference while on the stand. Even with the benefit of greater time and digital copy, the Court was similarly unable to find such a general reference.

[13] As explained by Dr. Garabrant, "[i]ssues such as confounding and effect modification are universal issues in complex, multivariate datasets. Epidemiology has worked out approaches to correctly recognize and, when possible, adjust for and deal with those issues. . . . [T]hose issues are not unique to ecologic datasets. They run through ecology, psychology, economics, human health studies. Those are just characteristics of large multivariable datasets." Tr. 1 at 50, ECF No. 105. While these principles and concepts may be universally applicable, Dr. Garabrant nevertheless managed to handily demonstrate how readily and unabashedly an analyst can misstep when applying them if he is unfamiliar with the subject matter under analysis.

Dr. Garabrant's misstep gives pause precisely because it was the result of a simple mathematical exercise; one that required mastery of little more than most are taught in high-school. Presumably, it is not Dr. Garabrant's competency at graphing that failed him, but his ability to sensibly interpret and analyze *ecological* data. As cautioned by Dr. Wing, "the further one goes from one's substantive knowledge, the more difficultly one would have even if the methods used, for example, in statistics were the same." Tr. 2 at 21, ECF No. 100.

Turning to Dr. Garabrant's broader critiques of the EPA Benchmark, we begin with the suggestion that the Benchmark is fundamentally flawed insofar as the EPA failed to account for effect modification. As explained by Dr. Garabrant, "[e]ffect modification occurs when the association between two factors is different depending on the presence or absence of a third factor. If, for example, the association between an exposure and an outcome is different for men than for women, sex modifies the relationship between the exposure and the outcome." ECF No. 90-1 at 3; *see also* Dictionary of Epidemiology (Miquel Porta ed., 6th ed. 2014) (defining "effect modification" as a "[v]ariation in the selected effect measure for the factor under study across levels of another factor"); *id.* (defining "effect modifier" as "[a] pre-exposure factor across whose levels the value of the effect measure of interest varies; [a] factor that biologically, clinically, socially, or otherwise alters the effects of another factor under study"). When effect modification is present, "[c]ombining the two groups to create a summary measure of association is meaningless: it is not true for men and it is not true for women." ECF No. 90-1 at 3. Thus, before a causal analysis moves on to continue potential confounders, it is essential to first assess whether effect modification is present. Dr. Garabrant, Tr. 1 at 14–15, ECF No. 105.

Returning to Dr. Garabrant's critique, indeed, the words "effect modification" cannot be found in the text of the Benchmark, suggesting to Dr. Garabrant that the EPA did no analysis of effect modification. In an effort to assess the presence or absence of effect modification, Dr. Garabrant turned to the underlying data and produced a series of figures purportedly showing the presence of effect modification. *See* Def.'s Exs. 31–36. In each table, the percent of sites with ephemeroptera present is represented along the y-axis, conductivity is represented along the x-axis, and blue, red, and green lines run across the graph as representations of low-, mid-, and upper-range values for a given potential effect modifier, respectively. Taking the example of pH

as a potential effect modifier (Def.'s Ex. 31), Dr. Garabrant explained that the effect of pH can be gleaned from looking at "the vertical distance between the green line and the blue line holding conductivity constant," with greater vertical distance suggesting greater likelihood of effect modification. Tr. 1 at 36–37, ECF No. 105.

Using these tables to thusly visualize the data, Dr. Garabrant reached the conclusion that effect modification was present with respect to pH (Tr.1 at 37, ECF No. 105 (discussing Def.'s Ex. 31: "[i]f pH is neutral to high, there is no relationship between conductivity and Ephermera. If pH is low, the insects are adversely affected. That's what effect modification looks like")), stream size (Tr. 1 at 39, ECF No. 105 (discussing Def.'s Ex. 32: "I think you have evidence here of effect modification . . . You are getting a different answer according to stream size. That's effect modification")), dissolved oxygen (Tr. 1 at 41, ECF No. 105 (discussing Def.'s Ex. 33)), iron (Tr. 1 at 43, ECF No. 105 (discussing Def.'s Ex. 34)), and manganese (Tr. 1 at 44–45, ECF No. 105 (discussing Def.'s Ex. 36)). Thus, the Court has one trained epidemiologist, with no formal background or experience in ecology, claiming that the EPA neglected to consider effect modification, and in so doing, missed the presence of several effect modifiers, thereby undermining the entirety of its causal analysis.

A second trained epidemiologist offered testimony on the same issue, but reached starkly different conclusions. Responding to Dr. Garabrant's analysis of effect modification and EPA's Benchmark, Plaintiffs' expert, Dr. Wing, cautioned that, "in order to make a decision about interaction or effect modification, it's first necessary to have some idea about the topic one is investigating because without that, one can make egregious mistakes about an analysis which can be done by someone who doesn't know anything about the topic but could result in essentially meaningless conclusions or actually conclusions that are misleading." Tr. 2 at 20, ECF No. 100.

He further stated the belief "that the issue of effect modification or interaction is one that should be made based on subjective knowledge in the area, and it's not one that's simply a statistical requirement or rule." Tr. 2 at 20, ECF No. 100.

Consistent with that fundamental reservation and despite his considerable epidemiological expertise, Dr. Wing was unable to agree with Dr. Garabrant's conclusion that effect modification is present in the dataset and yet left unaddressed by EPA. Looking, for example, at Dr. Garabrant's figure assessing dissolved oxygen as a potential effect modifier (Def.'s Ex. 33), Dr. Wing observed similar trends across low-, mid-, and upper-range dissolved oxygen levels. Tr. 2 at 45, ECF No. 100. According to Dr. Wing, the similarity of the trend, or slope, suggests an *absence* of effect modification. *Id.* Moreover, solely based on the graph relied upon by Dr. Garabrant, Dr. Wing explained that it was impossible to definitively assess effect modification because Dr. Garabrant neglected to include any information on sample size or precision (e.g., no slope estimates or standard error estimates are provided). Tr. 2 at 46, ECF No. 100. And so the opinion of a second epidemiologist without ecological training reaches not only a contrary conclusion about effect modification, but further identifies an analytical barrier to reliably interpreting the graphs relied upon by the first.

To that uncertain mix, Dr. Baker contributes his opinion on effect modification as an ecologist without formal epidemiological training. Like Dr. Wing, Dr. Baker similarly critiqued the absence of information on sample size or precision.[14] Dr. Baker further called attention to the

[14] On cross-examination, Dr. Baker was asked why he did not recreate plots, just as Dr. Garabrant had done, and calculate measures of significance necessary to interpret the plots created by Dr. Garabrant. Tr. 4 at 236–38, ECF No. 107. Like Dr. Garabrant, Dr. Baker did have access to the dataset and the expertise necessary to repeat the analysis. Since he could have repeated Dr. Garabrant's analysis to reveal error bars and since we find ourselves in an adversarial context, it is not unreasonable to ask why Dr. Baker did not do so.

That said, even in an adversarial context, it is arguably reasonable to expect that experts

fact that values were binned across the conductivity gradient by Dr. Garabrant in a manner that failed to control for sample size within each bin.[15] Most surprising, however, was Dr. Baker's testimony that the EPA *did* assess effect modification. According to Dr. Baker, the term "effect modification" is not commonly used in ecology. Tr. 4 at 239, ECF No. 107. Instead, ecologists commonly refer to "covariation" as a "catchall term" used for both confounding and effect

---

will proffer opinions in such a way that enables others to evaluate the significance and reliability of their work. As reminded by the current and past presidents of the International Epidemiological Association in the Forward to the Sixth Edition of the Dictionary of Epidemiology, "[t]he nature of science is not to reach consensus but to advance our knowledge by bringing conflicting ideas to critical examinations." Consistent with this interest in enabling critical examinations, in the context of science conducted purely for sake of advancing knowledge, experts are expected to provide sufficient information to enable peer review. This expectation of transparency ensures integrity and allows for meaningful discussions of what is known or unknown, and to what degree. It is not unreasonable for a court to have similar minimum expectations in the service of reaching reliable final judgments.

Thus, the better question might have been why did Dr. Garabrant *not* include indicators of significance in the analysis he conducted and presented to this Court. One can only imagine what dimensions of bias may be introduced to expert findings when moved from the theoretically unbiased universe of Science to the purposefully adversarial universe of courts. Without making judgments as to motive, reasonable possible inferences could be as troubling as thinking that Dr. Garabrant did not do so because he did not understand the interpretive significance of such indicators—calling the competency of his analysis more generally into question—or that he did not do so precisely because he did understand the interpretive significance—calling the credibility of his analysis more generally into question. Such musings could and should be rendered unnecessary merely by experts taking care to present findings to the Court with a degree of transparency that enables critical examination of conflicting ideas.

[15] Whatever conclusion is to be reached in light of the alleged obstacles to interpreting the significance of the plots, Dr. Baker added that conclusions regarding pH would be immaterial because EPA controlled for any effect modification occurring at pH<6 by truncating the data accordingly.

Though criticized, truncating data and sample selection appear to be accepted methodological approaches to control for confounding. *See e.g.* Virginia Tech, Pls.' Ex. 173 at PE 1692 ("This was accomplished by seeking study streams with attributes such as habitat quality that were as similar as possible to minimally-disturbed reference streams of the region. The design was intended to ensure that TDS, including its component ions, was the primary factor associated with biotic stress in these streams"); Dr. Wing, Tr. 2 at 32–33, ECF No. 100 (explaining the methodological validity of truncating data as a means of controlling for potential confounding effects).

modification. Tr. 4 at 191, ECF No. 107. Though never using the term, Dr. Baker remains assured that the EPA considered effect modification through alternate means. *Id.* at 191, 239.

Even standing independently, the rebuttal arguments offered by Dr. Wing and Dr. Baker arguably do enough to dispose of Dr. Garabrant's critique of the EPA Benchmark related to analysis of effect modification. Dr. Wing's testimony effectively draws the adequacy of Dr. Garabrant's expertise into question, and Dr. Baker's testimony demonstrates the importance of ecological expertise in reading and evaluating the EPA's work. Furthermore, Plaintiffs' expert testimony does not stand alone; it is accompanied by and consistent with the expertise and analysis of the EPA—an expert federal agency acting in its area of expertise.[16]

2. Analysis of Confounding

Dr. Garabrant further argued that the EPA performed an unreliable analysis of confounding, thereby rendering EPA's causal conclusions invalid. The Dictionary of Epidemiology defines "confounding" as

> [T]he distortion of a measure of the effect of an exposure on an outcome due to the association of the exposure with other factors that influence the occurrence of the outcome. Confounding occurs when all or part of the apparent association between the exposure and the outcome is in fact accounted for by other variables that affect the outcome and are not themselves affected by exposure."[17]

[16] As previously noted by the Court, in developing the Benchmark, the EPA's findings were subjected to review and comment by the Science Advisory Board. *Fola (Stillhouse)*, 2015 WL 362643, at *6. As noted by Dr. Garabrant, not only are there epidemiologists on the Science Advisory Board, "there are some very fine [epidemiologists]" serving in that capacity. Tr. 1 at 71, ECF No. 105. While offering constructive criticisms for further analysis, these epidemiologists evidently accepted the EPA's Benchmark as adequate, to say the least.

[17] In addition to this textbook definition, we might also keep in mind that, "confounding, like many concepts in science, is a basic concept, but it's not a settled concept, meaning that there are disagreements between epidemiologists, even in textbooks, about exactly how to define 'confounding.'" Dr. Wing, Tr. 2 at 25, ECF No. 100; *see also* Sharon Schwartz et al., Toward a Clarification of the Taxonomy of "Bias" in Epidemiology Textbooks, *Epidemiology* 26:2 (March 2015), Pls.' Ex. 170 ("We identified and reviewed 28 textbooks that met the inclusion criteria. Our review verified the norm of categorizing bias into confounding, selection bias, and information bias. All textbooks that included an organizing scheme (24/28) used these categories. However,

As argued by Dr. Garabrant, the EPA failed to adequately and reliably assess confounding, instead relying on an unverified and subjective methodology. Tr. 1 at 16–18, ECF No. 105.

Asked to comment on the validity of EPA's approach to analyzing confounding, Dr. Garabrant hestitated to say whether the approach was valid or not. Tr. 1 at 16, ECF No. 105 ("It's hard to say whether it is valid. I have never seen it used. I haven't seen any validation of it. I have never seen any test of this method to show that it works. So I would say it's not known whether it's reliable or not. EPA created it."). Beginning broadly, Dr. Garabrant called attention to the following paragraph from the Benchmark explaining the EPA's approach to confounding in its causal analysis:

> Weighing evidence for confounding factors differs from weighing evidence for causation. The causal assessment in Appendix A determines whether dissolved salts are an important cause of biological impairment in the region. This assessment of confounding accepts the result of the causal assessment and attempts to determine whether any of the known potential confounders interfere with estimating effects of conductivity to a significant degree.

EPA Benchmark at B-3. As explained by Dr. Garabrant, this paragraph reflects an analytical error on the part of the EPA akin to "putting the cart before the horse." Tr. 1 at 17, ECF No. 105 ("[T]he idea that you accept the result of causal assessment and then look at confounding is simply putting the cart before the horse. It's backwards."). In addition to this analytical error, Dr. Garabrant suggested that EPA's approach was relatively arbitrary and subjective. Tr. 1 at 16, ECF No. 105.

Dr. Garabrant then went on to explain that epidemiologists commonly rely on a relatively straightforward way to identify the presence of confounding effects: compare the results of a crude

---

only one textbook articulated the 2 elements of a consistent taxonomy—a feature that unites confounding, selection bias, and information bias and a feature that differentiates them. There was variation across textbooks as to how close they came to a clear description of the structure of their schema.").

analysis testing the association between conductivity and extirpation against the results of an adjusted analysis that introduces a potential confounding factor.[18] Tr. 1 at 18–19, ECF No. 105. Using the same dataset relied upon by the EPA, Dr. Garabrant performed precisely that analysis. *See* Def.'s Ex. 40. The results of that analysis are reproduced in the following table:

Variable	Adjusted parameter estimate of conductivity	Percent change in parameter estimate after adjustment
Dissolved Manganese	-2.13	81.94%
Sulfate	-3.16	22.85%
Total Magnesium	-4.77	18.67%
Total Orthophosphates	-4.69	17.17%
Hardness	-4.45	12.81%
Alkalinity	-4.42	12.23%
Watershed Square Kilometers	-4.41	12.04%
Dissolved Calcium	-3.47	11.87%
Total Selenium	-4.40	11.82%
Dissolved Orthophosphates	-4.40	11.72%
Total Manganese	-3.50	10.90%

Table 1: Confounders present in the EPA dataset. Adjusted conductivity parameter estimate and percent of change effected by each confounder. Crude conductivity (logarithm transformed) parameter estimate is -3.88. Def.'s Ex. 40.

Referring to this table, Dr. Garabrant offered testimony that any change in parameter estimate after adjusting for a given variable greater than 10% signals the presence of confounding. Tr. at 71–72. On cross-examination, however, Dr. Garabrant readily acknowledged that while "there is widespread agreement that more than fifty percent change is important," in the range of ten to twenty percent, judgments about confounding would depend on the analyst's background

[18] In contrast, Dr. Wing offered testimony that "there is not a set of rules for saying whether or not there's confounding." Tr. 2 at 27, ECF No. 100. Moreover, "different epidemiologists have somewhat different definitions of 'confounding'" such that "even if there were rules, they would have different rules." Dr. Wing, Tr. 2 at 27–28, ECF No. 100. Ultimately, according to Dr. Wing, "[you have to understand the substance of the topic" when conducting a causal analysis precisely because "[t]here's no generic analytic method which produces reliable science." Tr. 2 at 28, ECF No. 100. Here, it is perfectly reasonable to expect that the EPA possesses such expertise.

knowledge in the subject matter at issue. Tr. 1 at 72, ECF No. 105. As already mentioned, Dr. Garabrant is without precisely that background knowledge. Accordingly, the Court is left to conclude that Dr. Garabrant's ten percent threshold is itself arbitrary and unreliable.

Beyond analytical differences of opinion, the Court is further unmoved by Dr. Garabrant's analysis of confounding given the quality of underlying data and the nature of certain variables. First, as will be discussed at length in the next section, some of the variables analyzed by Dr. Garabrant for confounding are known to ecologists to have little to no relevance in the context of West Virginia streams impacted by alkaline mine drainage. (e.g., orthophosphates, *see* i*nfra* Section II.B.3.a). Second, and also discussed below, the database lacks a significant number of data points for some of these variables *unless* total and dissolved values are considered together. (e.g., magnesium, calcium, selenium, and manganese, *see infra* Section II.B.3.a).

Like Dr. Garabrant, this Court would be unable to set anything but an arbitrary threshold for recognizing potential confounding variables. Instead, the Court continues to rely on the expertise of ecologists and testimony assuring the Court that the EPA engaged a reasonable and verified analysis of confounding.[19]

3. Adequacy of the underlying data

[19] Furthermore, the Court notes that EPA's findings are corroborated by considerable peer-reviewed scientific literature. *See infra* Section II.C. Such studies, which will be reviewed below, reach corroborating conclusions regarding the causal relationship between ionic pollution and loss of freshwater macroinvertebrates despite reliance on unique methodologies and distinct datasets. Dr. Baker estimated that, across the published literature, roughly six to ten different statistical techniques have been relied upon to test that causal relationship. Tr. 3 at 123, ECF No. 106. Whatever technique is used, experts have consistently identified conductivity as the most likely cause of biological impairment. Similarly, Dr. Baker estimated that at least five different methods had been used across the published literature to analyze potential confounding factors. *Id.* at 124. Again, whatever the method used, experts have consistently ruled out potential confounders and identified conductivity as the most likely cause of biological impairment.

Dr. Garabrant offered two critiques of the data relied upon by EPA in developing the Benchmark. First, Dr. Garabrant highlighted missing data points (e.g., limited number of data points for dissolved calcium). Second, Dr. Garabrant criticized EPA's presentation of data and its analysis thereof as misleading.

a. "Missing" Data

With respect to allegations of fatally missing data, Dr. Garabrant prepared a table reporting the number and percent of missing data points for each variable missing greater than 50% of the possible data points. Def. Ex. 38 (reproduced below).

Variable	Number Missing (out of 2,210 total observations)	Percent Missing (out of 2,210 total observations)
Dissolved magnesium	2209	99.95%
Dissolved calcium	2201	99.59%
Dissolved manganese	2190	99.10%
Dissolved orthophosphates	2171	98.24%
Total orthophosphates	2170	98.19%
Dissolved selenium	1897	85.84%
Total selenium	1714	77.56%
Percent brush	1493	67.56%
Percent barren	1493	67.56%
Percent wetland	1493	67.56%
Percent agricultural	1493	67.56%
Percent urban	1493	67.56%
Percent woodland	1493	67.56%

Table 2: Number and percent of missing for each variable (Missing > 50% only). Def.'s Ex. 38.

According to the table, greater than 98% of the data points are missing for dissolved magnesium, dissolved calcium, dissolved manganese, dissolved orthophosphates, and total orthophosphates. Def.'s Ex. 38. Additionally, between 67% and 78% of the data was missing for various land cover variables, 85% of the data was missing for dissolved selenium and 77% missing for total selenium. According to Dr. Garabrant, these deficiencies in the dataset prevented the EPA from

meaningfully analyzing potential effects of these variables.[20] Tr. 1 at 24–25, ECF No. 105.

However, as an ecologist, Dr. Baker was not similarly troubled by the missing data. First, he explained that deficiencies found in dissolved magnesium, were well compensated for by data on total magnesium.[21] Tr. 4 at 199–201, ECF No. 107. As is the case with magnesium, greater than 50% of the sites in the WVDEP database had available data points for total calcium and total manganese. *Id.* Similarly, when the data points for both dissolved and total selenium are combined, greater than 50% of the sites had data on selenium levels.[22] *Id.*

With respect to alleged missing data on land cover categories, Dr. Baker was similarly untroubled. Dr. Baker testified that, as an ecologist, one would not be likely to consider any of these variables as potential confounding variables and would therefore likely ignore these categories as immaterial. *Id.* Finally, with respect to orthophosphates, Dr. Baker explained again that the lack of data would not trouble an ecologist in this context, because orthophosphates in high concentrations are associated with agricultural landscapes, which are generally not found near—much less coextensive with—mining areas in West Virginia. *Id.* at 201. Stated differently, the absence of land cover data or orthophosphate measurements would only trouble a reviewer to the extent that he did not have the necessary background familiarity with ecology and land use

[20] Surprisingly, Dr. Garabrant did not similarly conclude that these perceived data deficiencies prevent meaningful analysis of potential effect modification. Nearly half of Dr. Garabrant's likely effect modifiers also appear in his table listing variables with insufficient datapoints. *Compare* Def.'s Ex. 38 *and* Def.'s Ex. 40. For example, we see that there is only one data entry for dissolved magnesium among a total 2,210 observations, yet Dr. Garabrant concluded that dissolved magnesium is a likely effect modifier. Similarly, Dr. Garabrant identified dissolved manganese as the most likely effect modifier, despite there being only twenty data points for dissolved manganese among a total 2,210 observations.

[21] Dr. Baker represented that the dataset includes over 1,000 data points for Total Magnesium, Total Calcium, and Total Manganese. No testimony was offered to suggest different effects based on whether substances are present in ionized or non-ionized forms.

[22] Notwithstanding observations about the amount of selenium data available, all experts agreed that the EPA Benchmark recognized that given the limited available data, occurrence and effects of selenium should be investigated further.

patterns to independently recognize the insignificance of the variables.

b. "Hidden" Data

In addition to purportedly fatally missing data, Dr. Garabrant also criticized the EPA Benchmark for "hiding data." As explained by Dr. Garabrant, a series of tables provided in Appendix B (Tables B-8 (habitat), B-13 (Embeddedness), B-15 (pH), B-23 (stream size), B-25 (dissolved oxygen), B-28 (iron), B-29 (Aluminum), B-30 (Manganese)) all share a common flaw: failure to include significant chunks of data for conductivity levels between 200 μS/cm and 1500 μS/cm, instead, only showing data at extreme conductivity conditions (i.e., < 200 μS/cm and > 1,500 μS/cm). In some instances, these tables fail to include significant mid-range data for not only conductivity, but also for the variable of interest (e.g., mid-range iron data was not included in Table B-28). Dr. Garabrant recognized and critiqued the absence of mid-range data, and was unable to provide any methodological or analytical justification for its absence.

Making good use of database access, Dr. Garabrant recreated the suspect tables to include mid-range data (*see* Def.'s Exs. 39, 43–49) and then continued to plot the data represented in each table (*see* Def.'s Exs. 30–36). Based on this information, Dr. Garabrant reached two conclusions: (1) effect modification is present with respect to each co-variate represented; and (2) there is a consistent absence of conductivity effects regardless of co-variate levels until conductivity reaches 1200 μS/cm to 1500 μS/cm. These conclusions suggested to Dr. Garabrant that there are serious flaws in the EPA's analysis. Tr. 1 at 33–46, ECF No. 105. In turn, Dr. Garabrant's only explanation for how the Benchmark reached publication despite hidden data became the suggestion that the SAB and peer reviewers would not have had access to the dataset and the ability to perform the analysis he did. Tr. 1 at 48–49, ECF No. 105.

Through the testimony of Dr. Baker and Dr. Wing, Plaintiffs supplied convincing rebuttals

to each criticism related to "hidden data" raised by Dr. Garabrant. Not only are these rebuttal arguments convincing, but to some extent, they also highlight the drawbacks of data analysis performed with relatively limited understanding of the subject matter being analyzed. First, Dr. Baker explained a methodological reason EPA did not include mid-range data in the tables: given the nature of the data, mid-range values were irrelevant to answering the question asked. Tr. 4 at 203–04, 207–11, ECF No. 107. The EPA dataset was qualitatively limited (or coarse) in that the data captured presence or absence of mayflies, but not information on abundance or variety. The category "mayflies" includes a variety of discrete species, each with particular sensitivity to conductivity. Some mayflies are uniquely sensitive, while others are uniquely tolerant to conductivity. As established in the then published literature, *all* mayflies—sensitive and tolerant alike—can be expected to have a negative response to conductivity levels in excess of 1,500 μS/cm. Stated differently, until conductivity exceeds 1,500 μS/cm, the available data would likely show *some* mayfly present.[23] If *some* mayfly is present, however, that does not necessarily tell us anything about abundance (i.e., only one bug could be present) or variety (i.e., only one species of mayfly could be present). Accordingly, if the question is whether or not mayflies may be present regardless of co-variate influences, the data only allow us to answer that question if we look to the extreme conductivity ranges (i.e., lowest conductivities where we would expect even the most sensitive mayflies to be present and the highest conductivities where we would expect even the most tolerant mayflies to be absent).

Given Dr. Baker's more nuanced explanation of the import of the data shown and the data

[23] As explained directly by Dr. Baker, "[t]he criteria the EPA was using here was to find any site with a mayfly. Now because the mayfly is an order, there are many families with different characteristics and also genera within those families, and we know that at least a few of them can tolerate exceptionally high conductivities. Therefore, it would not be surprising at all to find mayflies, a mayfly, at nearly all of the sites between 200 and 1,500 μS/cm." Tr. 4 at 203–04, ECF No. 107.

not shown, Dr. Garabrant's criticism of hidden data does little to impugn the work of EPA scientists, instead illustrating Dr. Wing's point that data analysis is not the same as the interpretation of data. While an epidemiologist may be qualified to run data analyses on any dataset, it should not be assumed that an epidemiologist is necessarily otherwise qualified to interpret the results of that analysis.

Second, Dr. Baker tailored his interpretation of the data presented according to the function of the tables in the overall analysis. Importantly, these tables were not relied upon by the EPA to identify threshold effects; these tables were introduced to explain EPA's confounding analysis. Presence or absence of some mayfly genera without any data on abundance or variety explains very little that would help to identify a conductivity threshold at which the most sensitive macroinvertebrates suffer extirpation.

4. Inter-state differences in species sensitivity

Dr. Garabrant's observations about the differences between XC95 values in West Virginia and in Kentucky are among his most immediately compelling observations. The Benchmark's XC95 values report genera-specific response thresholds at which you can expect 95% of freshwater macroinveterates to tolerate conductivity levels. While it is not surprising to expect different genera to have different response thresholds, Dr. Garabrant made the troubling observation that there are variances in the response thresholds *within* genera based on whether the data was sourced from West Virginia or Kentucky. For example, cross comparison of tables reveals the following differences between response thresholds for genera in West Virginia and Kentucky, among others:

Genera	West Virginia	Kentucky
Diphetor	632	190
Oulimnius	>2,791	320
Pycnopsyche	295	>775
Wormaldia	>1,553	235
Dolophilodes	>863	270
Oulimnius	>2,791	320
Ablabesmyia	>11,646	>1,410
Lepidostoma	~121	149

Table 3: XC95 thresholds identified in EPA's Benchmark for selected genera under the West Virginia and Kentucky datasets.

It is easy to share Dr. Garabrant's shock upon noticing that different genera apparently have different response thresholds depending on whether they are observed in one state or another.[24] Surely biological responses should not vary according to political boundaries; indeed, as aptly stated by Dr. Garabrant, "bugs don't know where they live." Tr. 1 at 29, ECF No. 105.

Though Dr. Garabrant could only imagine such discrepancies suggested fatal methodological and analytical flaws, Dr. Baker readily offered sound explanations rooted in the nature of data analysis and data collection. Tr. 4 at 211–14, ECF No. 107. First, whatever the differences in XC95 values of a few genera, across the entire datasets, the XC95 are very well correlated.[25] *Id.* at 211–12. Second, the West Virginia dataset includes several thousand samples; the Kentucky dataset includes roughly two hundred. *Id.* at 213–14. All else being equal, based on sample size alone, the Kentucky data would have a larger possibility of error and the West Virginia database would be more reliable. *Id.* In addition to differences in the quantity of data, the quality of the data for each state is unique. *Id.* at 214–15. In Kentucky, sampling protocol

[24] It is worth noting that Plaintiffs are not complaining of conductivity levels that mildly exceed EPA's Benchmark of 300 µS/cm. With few exceptions, the conditions at issue here exceed nearly all the XC95 threshold values from either state.

[25] Dr. Baker further noted that the particular taxa discussed by Dr. Garabrant "represent the most extreme outliers, so they are an example of just taking the [worst] cases to make a comparison." Tr. 4 at 212, ECF No. 107.

directs that *all* bugs collected in the sample be counted. *Id.* In West Virginia, the sampling protocol directs that only a subset of the total sample be counted. *Id.*

Given these differences in collection methods and database size, one would expect that the EPA would identify two different and loosely associated benchmarks. Yet the methodology used by the EPA identified remarkably similar benchmark values notwithstanding species-specific differences.[26] Dr. Garabrant correctly concludes that the datasets are imperfect—individually and relatively. So far as this Court understands ecological study, datasets are invariably imperfect. Notwithstanding perennial deficiencies in information, it remains the task of the scientist to distill reliable (and in the regulatory context, actionable) results. Based on Dr. Baker's explanation of the distorting effect of the quality and quantity of data in West Virginia as opposed to Kentucky, the Court remains confident that the EPA Benchmark presents reliable findings based on the information available. This is particularly the case with respect to the EPA's analysis and conclusions based on the WVDEP database. Tr. 4 at 216, ECF No. 107 ("Overall, I would expect the West Virginia dataset to be a little bit more precise given the nature of the data size.").

C. Scholarly Publications and Expert Opinions

In addition to EPA's Benchmark, Plaintiffs further relied on a seemingly ever-growing collection of published, peer-reviewed journal articles addressing the connection between conductivity and impairment in Appalachian streams. Through the testimony of experts, the Court was introduced to myriad peer-reviewed articles. In revisiting that collection of articles below, note the complete absence of peer-reviewed scientific articles to the contrary.[27] Tr. 2 at 95,

[26] Dr. Baker offered testimony explaining that the XC95 values in both states were closely correlated even in the aggregate and in an analog scale. Tr. 4 at 227, ECF No. 107.

[27] In collaboration with colleagues, Dr. Menzie developed a paper titled "A Cautionary

ECF No. 100. Instead, the scientific community repeatedly reaches and reports the same conclusion despite the use of multiple methodologies relying on a variety of datasets and conducted by a range of expert scientists.[28] Given that growing and consistent body of scientific study, it is not surprising that Dr. Palmer is of the opinion that "there is a strong relationship and evidence of causation between high conductivity and impairment" in central Appalachian streams impacted by alkaline mine drainage. Tr. 2 at 94, ECF No. 100 ("I have at this point absolutely no doubt. There are so many studies that have been done, using very different methods and very different places that have all reached the same conclusion."); *see also* Tr. 3 at 125–26 (Dr. Baker explains that the relationship between elevated conductivity and biological impairment is very strongly supported, to the point that he would sooner consider it a fact of science than a theory).

The scientific literature concerning the relationship between conductivity and impairment likely began in earnest in 2003 with publication of an Environmental Impact Study (EIS) of mountaintop mining valley fills. Tr. 2 at 99, ECF No. 100. Authors of the EIS examined changes in water chemistry and biological assemblages, finding increased concentrations of sulfates and dissolved solids, increased specific conductance, and a coincident decrease in sensitive taxa in impacted streams. Tr. 2 at 99–100, ECF No. 100. Other early publications included a paper by

---

Note About Deriving Causal Relationships in Water Quality Benchmarks from Field Observations Data: A Case Study in West Virginia Headwaters." That paper includes some of the same material, tables, and conclusions offered as testimony before this Court, on this and previous occasions. Tr. 4 at 89, ECF No. 107. Development for publication was supported and funded by Fola, Alpha, and Rio Tinto. *Id.* at 90–91. Dr. Menzie submitted the paper for publication to *Environmental Science and Technology*, a journal which has previously published Dr. Menzie's work and the journal that published a series of articles by Cormier and Suter ostensibly comprising the EPA's Benchmark. *Id.* at 89–90. "A Cautionary Note" was not accepted for publication.

[28] As relayed by Dr. Palmer, "[o]ne of the things we're taught very early on as a scientist, that the strongest form of inference you can make is if you have multiple ways to reach the same conclusion; if you use multiple methods, different kinds of experiments, observations, and particularly if different people do this work so you can eliminate potential methodological differences that individuals might impart." Tr. 2 at 96, ECF No. 100.

Kennedy et al. the following year, relaying the finding that exposure to elevated conductivity levels resulted in loss of organisms (Joint Ex. 9; Tr. 2 at 100, ECF No. 100), and a 2005 publication by Hartman examining the relationship between conductivity and mayfly richness. Tr. 2 at 100, ECF No. 100.

In 2008, Gregory Pond et al. published a paper in the Journal of North American Benthological Society, titled "Downstream effects of mountaintop coal mining: comparing biological conditions using family- and genus-level macroinvertebrate bioassessment tools."[29] Joint Ex. 13. In the underlying study, the authors conducted field sampling in order to analyze differences in water chemistry and macroinvertebrate assemblages at mined and unmined sites. Tr. 2 at 104. Pond et al., concluded that there was strong evidence of a causal relationship between conductivity and biological impairment. *Id.* at 104 ("Our results indicate that [mountaintop removal mining] is strongly related to downstream biological impairment, whether raw taxonomic data, individual metrics that represent important components of the macroinvertebrate assemblage, or [multimetric indexes] are considered. The severity of the impairment rises to the level of violation of water-quality standards (WQS) when states use biological data to interpret narrative standards."). Moreover, the authors particularly noted that mayflies were especially sensitive to changes in water chemistry. *Id.* at 104.

Furthermore, Pond et al. calculated correlation coefficients of "[GLIMPSS] and [WVSCI] and genus- and family-level nonmetric multidimensional scaling (NMS) axis scores [verses] a truncated list of environmental variables," including conductivity, embeddedness scores, sediment deposition scores, and total rapid biological protocol ("RBP") habitat scores.[30] *See* Table 5, Pls.'

[29] Authors include Gregory J. Pond, Margaret E. Passmore, Frank A. Borsuk, Lou Reynolds, and Carole J. Rose.

[30] The EPA's rapid bioassessment protocol is a multimetric tool for scoring habitat quality

Ex. 173 at JE0196. The authors found statistically significant correlations between the metrics and the total RBP habitat score (GLIMPSS 0.38; WVSCI 0.43), but not the embeddedness scores (GLIMPSS 0.23; WVSCI 0.22) or the sediment deposition scores (GLIMPSS 0.20; WVSCI 0.28). *Id.* The correlation coefficient for conductivity was almost two-fold the other values (GLIMPSS –0.91; WVSCI –0.80). *Id.* These findings support the conclusion that '[w]ater quality structured benthic communities more than habitat quality." Pls.' Ex. 173 at JE0198.

Though relatively insignificant, Pond 2008 did find some positive correlation between habitat quality and aquatic life; however, subsequent studies have made further efforts to parse the difference between habitat influenced effects and the effects of water chemistry. *See e.g.* Pls.' Ex. 173 at PE 1537 ("This suggests that degradation of water quality and the resultant increases in specific conductivity, component ions, and trace metals limit aquatic life regardless of habitat quality."); Virginia Tech, Pls.' Ex. 173 at PE 1703 ("Nonetheless, the extensive effort undertaken to locate test sites with abiotic conditions comparable to those of reference sites was successful in minimizing biotic influence from non-TDS [total dissolved solids] stressors, including poor habitat quality. This was an important step toward defining TDS sensitivity . . .")[31]; Pond 2014 ("Habitat can be a limiting factor, but by design, we removed significant habitat degradation factors by selecting sample reaches with relatively good habitat and intact riparian vegetation at reference and VF sites . . ."); *id.* ("Overall, biological variation was strongly correlated with water chemistry and less by reach-scale habitat and landscape conditions. Since ion concentrations explained the greatest amount of biological impacts and were the most altered (compared to reference), this suggests that recovery is potentially hindered by ions, even in forested reaches long

---

based on the aggregation of scores assigned to ten different variables. Total RBP scores are grouped into four different categories, or levels: optimal, suboptimal, marginal, and poor.

[31] "TDS stressors" would include the sort of ionic pollution alleged in this case. Similarly, an assessment of "TDS sensitivity" would include analysis of ionic pollution sensitivities.

after reclamation."); Hitt et al, Joint Ex. 8, Pls.' Ex. 173 at JE0129 (regarding impacts to fish assemblages, the authors noted that "[o]bserved effects of [mountaintop removal mining] could not be explained by changes in physical habitat conditions").

The following year, Pond published a second article, "Patterns of Ephemeroptera taxa loss in Appalachian headwater streams," in *Hydrobiologia*. Pls.' Ex. 131. As was the case with his earlier work, Pond again relied on experimental fieldwork, but he conducted unique fieldwork in a different area. *Id.* In this second article, Pond compared mayfly assemblages at some ninety-two sites in Kentucky, focusing on taxa richness (i.e., "the number of different groups of mayflies," Tr. 2 at 105, ECF No. 100) and relative abundance. In so doing, Pond discovered that both mayfly richness and relative abundance were significantly higher at reference sites and both were significantly lower at mined sites. Tr. 2 at 105–06, ECF No. 100. Furthermore, consistent with earlier analyses, Pond reported that "[r]elative mayfly abundance was most strongly correlated to specific conductance (r = 0.72) compared to total habitat score (r = 0.59)." Pls.' Ex. 173 at PE1526; *id.* at PE1536 ("Analyses from WV mining areas (Hartman et al., 2005; Merricks et al., 2007; Pond et al. 2008) indicated that the decline in mayflies from mountaintop mining correlates most strongly to specific conductance.").

In the same year that Pond 2010 was published, *Science Policy Forum* published an article co-authored by Dr. Palmer and titled "Mountaintop Mining Consequences."[32] Pls.' Ex. 133. Here, rather than experimental fieldwork, the authors relied on data from a variety of sources, including data from the WVDEP database. Tr. 2 at 106, ECF No. 100. Through that data, the authors again examined the relationship between water chemistry and mining activities. *Id.* at 106.

[32] Additional authors include E.S. Bernhardt, W.H. Schlesinger, K.N. Eshleman, E. Foufoula-Georgiou, M.S. Hendryx, A.D. Lemly, G.E. Likens, O.L. Loucks, M.E. Power, and P.R. Wilcock.

In so doing, the authors again observed that mining contributed to poor water chemistry, particularly marked by elevated conductivity levels, and that significant declines in macroinvertebrate taxa resulted. *Id*.

The following year, the *Journal of the North American Benthological Society* published a paper by Eric Merriam et al., titled "Additive effects of mining and residential development on stream conditions in a central Appalachian watershed." Joint Ex. 11. As explained by Dr. Palmer, Merriam et al., examined the combined effects of streams impacted by mining as compared to streams otherwise impacted by development, finding that mining impacts do contribute to changes in macroinvertebrate community structure. Tr. 2 at 108–09, ECF No. 100. These changes in community structure would appear to have been more closely related to changes in water chemistry as compared to changes in habitat. Pls.' Ex. 173 at JE0173–74 ("We found significant effects of mining on in-stream conditions. Increased levels of mining resulting in poorer water quality, primarily through increases in specific conductance and associated dissolved chemical constituents. . . . Mining had no measurable effect on habitat complexity or quality."). Thus, relying on unique data and methodology the authors were able to conclude as follows:

> Our results are similar to those of recent studies that have identified changes in water quality to be the dominant stressor in mined systems (Fulk et al. 2003, Freund and Petty 2007, Pond et al. 2008, Petty et al. 2010, Pond 2010). Increased specific conductance is consistently the dominant stressor in streams affected by mountaintop removal mining in southern West Virginia (Hartman et al. 2005, Merricks et al. 2007, Pond et al. 2008). . . . Furthermore, increased specific conductance is a consistnely important predictor of ecological condition in these systems . . . Our results corroborate those of numerous studies in which Ephemeroptera was identified as one of the most sensitive taxa to increases in ionic strength associated with large-scale surface mining in the Central Appalachian region.

Pls.' Ex. 173 at JE0174.

This brings us to publication of EPA's Benchmark. As evident by the foregoing discussion, by the time the EPA published the Benchmark, scientific literature on the subject was already well developed, and according to Dr. Palmer, had already established a likely relationship between conductivity and impairment in Appalachian streams impacted by alkaline mine drainage. Tr. 2 at 110, ECF No. 100. Nevertheless, studies examining the probable relationship between mining, high conductivity, and impairment continued to reach publication in peer-reviewed scientific journals.

In 2011, Dr. Bernhardt and Palmer published an article titled "The environmental costs of mountaintop mining valley fill operations for aquatic ecosystems of the Central Appalachians" in the *Annals of the New York Academy of Sciences*. Joint Ex. 1; Tr. 2 at 117, ECF No. 100. There, the authors concluded that there was a significant relationship between mining activities and changes in the chemical composition of streams below mining. Tr. 2 at 117, ECF No. 100. Such changes were strongly associated with biological impairment of those streams. *Id.* at 117. Particularly, the authors explained that "[a]ll available data show that it becomes increasingly unlikely to find an unimpaired aquatic benthic community as conductivity increases." Pls.' Ex. 173 at JE0010. Elaborating on the same point, the article goes on to say that:

> Whether or not individual component ions within mining-derived runoff reach streamwater concentrations that are individually lethal or toxic to aquatic life, the cumulative effect of elevated concentrations of multiple contaminants is clearly associated with a substantial reduction in water quality and biological integrity in streams and rivers below mine sites. ***All research to date indicates that conductivity is a robust measure of the cumulative or additive impacts of the elevated concentrations of multiple chemical stressors from mine sites that lead to biological impairment of streams.***

Pls.' Ex. 173 at JE0014 (emphasis added).

Dr. Lindberg next joined Dr. Bernhardt on a paper published in the *Proceedings of the National Academy of Sciences* titled "Cumulative impacts of mountaintop mining on an

Appalachian watershed."[33] Pls.' Ex. 136. In this study, the authors "document the cumulative impact of more than 100 mining discharge outlets and approximately 28 km^2 of active and reclaimed surface coal mines on the Upper Mud River of West Virginia." Pls.' Ex. 173 at PE1759. In so doing, they observed that "[a]ll tributaries draining mountaintop-mining-impacted catchments were characterized by high conductivity and increased sulfate concentration." *Id.* More broadly, the unique approach taken in this paper established the cumulative impacts of mining in a watershed, with conductivity, sulfates, and selenium all significantly increasing with increased mining. Tr. 2 at 118–19, ECF No. 100; Pls.' Ex. 173 at PE1763 ("Our synoptic survey approach conclusively demonstrates that the observed increases in conductivity and [selenium] concentration can be attributed directly to the areal extent of surface coal mining occurring in the watershed.").

Still in 2011, another study authored by Dr. Pond reached publication in *Hydrobiologia*: "Biodiversity loss in Appalachian headwater streams (Kentucky, USA): Plecoptera and Trichoptera communities." Pls.' Ex. 137. Here, Dr. Pond again documented the effects of mining and residential land use disturbances on macroinvertebrates, particularly stonefly (Plecoptera) and caddisfly (Trichoptera) assemblages. Tr. 2 at 119, ECF No. 100. Dr. Pond found not only extirpation of these genera associated with mining disturbances, but further remarked that habitat factors could not explain the observed impacts. Tr. 2 at 119–20, ECF No. 100; Pls.' Ex. 173 at PE1775–76 ("no habitat factors were significantly correlated with relative abundance metrics").

In 2013, Pond published yet another co-authored paper on the subject; this time in *Environmental Monitoring and Assessment* and titled "Calibration and validation of a regionally and seasonally stratified macroinvertebrate index for West Virginia wadeable streams." Pls.' Ex.

[33] The *Proceedings of the National Academy of Science* is regarded as a rigorously peer-reviewed scientific journal. Tr. 2 at 119.

138. The article describes "the development, validation, and application of a geographically- and seasonally partitioned genus-level index of most probable stream status (GLIMPSS) for West Virginia wadeable streams." Pls.' Ex. 173 at PE1786. Importantly, the genus-level index developed therein proved to be a more reliable predictor of stream quality than its family-level counterpart, WVSCI. Tr. 2 at 121, ECF No. 100 ("pointing out that a genus-level index is much more appropriate to use because the family-level index is not adequately sensitive [] because it lumps genera that have very different tolerance levels."). These findings would suggest that WVDEP reporting of streams impaired according to WVSCI score is under-inclusive. Pls.' Ex. 173 at PE1803 ("Overall, GLIMPSS rate more than twice the number of sites as 'severely degraded' compared with WVSCI across all strata.").

Adding to the variety of methods used to address the question, Bernhardt et al., later published "How Many Mountains" in *Environmental Science and Technology*. Joint Ex. 2. In an attempt to further assess the extent of pollution resulting from mining, the authors "mapped surface mining from 1976 to 2005 for a 19,581 km^2 area of southern West Virginia and linked these maps with water quality and biological data for 223 streams." Pls.' Ex. 173 at JE0020. In so doing, they observed that the amount of mining in an area was highly correlated to increased conductivity in area streams. Pls.' Ex. 173 at JE0020 ("The extent of surface mining within catchments is highly correlated with the ionic strength and sulfate concentrations of receiving streams."). Then, relying on generalized additive models, the authors identified the following thresholds at which amount of watershed mining, stream ionic strength, or sulfate concentrations render impairment likely: "We find this threshold is reached once surface coal mine occupy >5.4% of their contributing watershed area, ionic strength exceeds 308 $\mu S/cm^{-1}$, or sulfate concentrations exceed 50 mg/L^{-1}." *Id.* Though the authors relied on different data and a unique methodology,

they nonetheless arrived at a conductivity threshold remarkably similar to the threshold identified by the EPA. Further consistent with the Benchmark, the authors did so only after controlling for the potential effects of habitat. Pls.' Ex. 173 at JE0022 ("Finally, [general additive models] allowed us to model the stressor-response relationship after controlling the effect of instream habitat quality, a variable that influences community metrics independently of catchment mining and stream chemistry.").

In 2013, Dr. James Kunz introduced yet another novel methodological approach and reached consistent results. Kunz et al., *Use of Reconstituted Waters to Evaluate Effects of Elevated Major Ions Associated with Mountaintop Coal Mining on Freshwater Invertebrates*, 32 Envtl. Toxicology & Chemistry 2826 (2013), Joint Ex. 10. The authors of this paper exposed selected freshwater organisms to mixtures with different ionic strengths in a laboratory in order to perform toxicity testing. Particularly of note, the ionic composition of some the reconstituted mixtures was representative of the ionic composition of central Appalachian waters impacted by alkaline mine drainage from mountaintop removal and valley fills:

> Two of the reconstituted waters had ionic compositions representative of alkaline mine drainage associated with mountaintop removal and valley fill-impacted streams (Winding Shoals and Boardtree, with elevated Mg, Ca, K, SO^4, HCO^3), and a third reconstituted water had an ionic composition repreentative of neutralized mine drainage (Upper Dempsey, with elevated Na, K, SO^4, and HCO^3). The waters with similar conductivites but, with different ionic compositions had different effects on the test organisms. The Winding Shoals and Boardtree reconstituted waters were consistently toxic to the mussel, the amphipod, and the mayfly.

Pls.' Ex. 173 at JE0152; Tr. 2 at 137, ECF No. 100. Furthermore, through laboratory analysis, Kunz et al., identified toxic impacts to *Centroptilum*, a mayfly, between 800 and 1,300 μS/cm; remarkably consistent with the Benchmark value of 1,092 μS/cm derived from statistical analysis of the WVDEP database. Joint Ex. 17 at D-3.

Next, the scientific literature returns to the authors of EPA's Benchmark, Drs. Susan Cormier and Glenn Suter, and the subsequent publication of several component sections of the Benchmark in the peer-reviewed scientific journal *Environmental Toxicology and Chemistry*.[34] In one of these component articles, Cormier and Suter analyzed six characteristics of causation: co-occurrence, preceding causation, interaction, alteration, sufficiency, and time order, finding all but one strongly supported the causal relationship, with no evidence available for the outstanding characteristic.[35] Susan M. Cormier et al., *Assessing Causation of the Extirpation of Stream Macroinvertebrates by a Mixture of Ions,* 32 Envtl. Toxicology & Chemistry 277 (2013), Joint Ex. 4. Cormier and Suter further found that "[t]he conductivity at mined sites is 10 to 50 times greater than at unmined sties. The source of increased conductivity is independently corroborated and consistent." Pls.' Ex. 173 at JE0083. Relying on multiple lines of evidence, including not only their own findings in developing the Benchmark, but also based on the collection of research available in the published literature, Cormier and Suter conducted a formal causal analysis linking high conductivity and extirpation of sensitive macroinvertebrates in central Appalachian streams:

> Through this assessment, the authors found that a mixture containing the ions

---

[34] *See* Cormier et al., *Derivation of a Benchmark for Freshwater Ionic Strength, supra* note 5; Cormier et al., *A Method for Assessing Causation, supra* note 5; Cormier et al., *Assessing Causation of the Extirpation of Stream Macroinvertebrates by a Mixture of Ions, supra* note 5; Cormier et al., *Relationship of Land Use and Elevated Ionic Strength in Appalachia Watersheds, supra* note 5; Cormier et al., *A Method for Deriving Water-Quality Benchmarks Using Field Data, supra* note 5; Suter et al., *A Method for Assessing the Potential for Confounding Applied to Ionic Strength in Central Appalachian Streams, supra* note 5.

[35] Of note, time order was the only factor that did not strongly support the causal relationship. Instead, Cormier and Suter found that they lacked the necessary evidence to assess whether time order supported the causal relationship, and therefore scored time order as providing "no evidence" either way. Pls. Ex. 173 (explaining that the authors "could not obtain conductivity and biological survey data collected before and after construction of a valley fill or release of ion-rich effluents from other sources. Hence, this characteristic of causation is scored as no evidence."). Though Cormier and Suter were unable to obtain the necessary information to assess the significance of time order in assessing the causal relationship, the record here is replete with such information and evidence.

> [calcium, magnesium, bicarbonate, and sulfate], as measured by conductivity, is a common cause of extirpation of aquatic macroinvertebrates in Appalachia where surface coal mining is prevalent. The mixture of ions is implicated as the cause rather than any individual constituent of the mixture. The authors also expect that ionic concentrations sufficient to cause extirpations would occur with a similar salt mixture containing predominantly [bicarbonate, sulfate, calcium, and magnesium] in other regions with naturally low conductivity.

Pls.' Ex. 173 at JE0080.

Cormier and Suter independently published their confounding factor analysis as well. Cormier & Suter, *A Method for Assessing the Potential for Confounding Applied to Ionic Strength in Central Appalachian Streams*, 32(2) Envtl. Toxicology and Chemistry 288 (2013), Pls.' Ex. 139. Using a weight-of-evidence approach, that analysis considers twelve potential confounders: habitat, organic enrichment, nutrients, deposited sediments, pH, selenium, temperature, lack of headwaters, catchment area, settling ponds, dissolved oxygen, and metals. Pls.' Ex. 173 at JE0091. By adapting principles of epidemiology to the applied study of multivariate ecological field data, Cormier and Suter examine and methodically eliminate each potential confounder. Particularly of note, Cormier and Suter considered and rejected the idea that embeddedness or the presence of upstream ponds confounds the relationship between conductivity and impairment. Pls.' Ex. 173 at JE0094 ("No evidence supported embeddedness as a factor"); *Id.* at JE0096 ("The weight of evidence for confounding from ponds is uniformly negative, so we conclude that the presence of ponds has little or no effect on invertebrate response to conductivity.").

More recently, in 2014, Drs. Pond, Margaret Passmore, Kelly Krock, and Jennifer Fulton—all with the EPA—along with Nancy Pointon, John Felbinger, Craig Walker, and Whitney Nash—colleagues from the OSMRE—published a peer-reviewed scientific article in *Environmental Management* finding, among other conclusions, that the vast majority of streams adjacent to reclaimed mine sites with valley fills were still impaired eleven to thirty-three years

after reclamation. Pond et al., *Long-Term Impacts on Macroinvertebrates Downstream of Reclaimed Mountaintop Mining Valley Fills in Central Appalachia*, 54(4) Envtl. Mgmt. 919 (October 2014), Pls.' Ex. 141, Tr. 2 at 129–33 ("Although these [valley fills] were constructed pursuant to permits and regulatory programs that have as their stated goals that (1) mined land be reclaimed and restored to its original use or a use of higher value, and (2) mining does not cause or contribute to violations of water quality standards, we found sustained ecological damage in headwater streams draining [valley fills] long after reclamation was completed"). The article explains that researchers "found that known sensitive taxa such as the mayflies *Ephemerella* and *Epeorus* and the caddisfly *Neophylax* were found at 100% of the reference sites but were absent from 12 of 15 (~80%) of the [valley fill] sites." Pls.' Ex. 173 at PE1832. As explained by Dr. Palmer, this most recent article by Dr. Pond and colleagues, based on a natural experiment, showed that the likely explanation for finding unexpected organisms in high conductivity waters is that the organisms are drifting into these areas. Tr. 2 at 130, ECF No. 100.

Furthermore, it is also of note that in selecting reference streams for this most recent study, Dr. Pond and his collaborators selected sites with comparable temperature and habitat regimes to the mined sites. Pls.' Ex. 173 at PE1827 ("Local reference streams were sighted in close proximity (range .75 to 10.5 km) to paired [valley fills] . . . and had similar catchment areas, forest types, and base geology"). That methodological approach had the effect of eliminating temperature and habitat scores as potential confounding factors. Pls.' Ex. 173 at PE1836 ("Habitat can be a limiting factor, but by design, we removed significant habitat degradation factors by selecting sample reaches with relatively good habitat and intact riparian vegetation at reference and [valley fill] sites"). The experiment also relied on reference sites that "were not pristine, as their catchments frequently had poorly maintained roads and culverts, utility right-of-ways, gas wells, or

underground mining that did not discharge to the watershed," thereby further eliminating potentially confounding factors. Pls.' Ex. 173 at PE1827.

On the basis of this outstanding collection of peer-reviewed studies, the Court finds that the link between surface mining and biological impairment of downstream waters has been sufficiently—if not definitively—established in the scientific literature. "There's field data. There's lab data. There's observational data. There's field experimental data. There's toxicity testing." Tr. 2 at 141, ECF No. 100. Through myriad lines of evidence, researchers have reached the same general causation conclusion, without a single peer-reviewed publication reporting contrary findings. In Dr. Palmer's expert opinion, there is no remaining doubt on the question of general causation, leaving only surprise that researchers are continuing to study the question. *Id.* at 141 ("I would say there's no doubt. What surprised me is that the studies continue to go on. . . . because it's been so well-established.").

III. Specific Causation

Having met the burden of establishing the general principle that high conductivity levels in streams, caused by alkaline mine drainage, lead to biological impairment, Plaintiffs' must next establish by a preponderance of the evidence that high conductivity levels, caused by Defendant's mine discharges, are causing or materially contributing to biological impairment in this particular instance.[36]

A. Stream Conditions

[36] *See* Sanne H. Knudsen, *The Long-Term Tort: In Search of a New Causation Framework for Natural Resource Damages*, 109 Nw. U. L. Rev. 475, 532 (Winter 2004) ("In the oil spill context, proving general causation would require showing that exposure to oil is capable of causing the alleged injury—e.g., disrupting the reproductive capacity of sea otters. Proving specific causation would require showing that sea otters were exposed to oil in doses capable of causing injury and that the oil came from the defendant's release.")

As introduced above, this case concerns discharges from three surface mines operated by Defendant: (1) Fola Surface Mine No. 2, discharging into Road Fork; (2) Fola Surface Mine No. 4A, discharging into Right Fork; and (3) Fola Surface Mine No. 6, discharging into Cogar Hollow. Largely through the testimony of Drs. Palmer, Swan, and Menzie, the Court heard detailed information on the each of the associated streams, spanning from before Defendant's mining operations began through to current conditions. Factual findings relating to pre-mining and current conditions for each surface mine are provided below. As demonstrated by that evidence, these streams have endured a pattern of increasing conductivity levels, increasing sulfates, and declining WVSCI scores.

1. Fola Surface Mine No. 2 – Road Fork

Fola Surface Mine No. 2 is regulated under WV/NPDES Permit No. WV1013840 and West Virginia Surface Mining Permit S201293. Stipulation ¶¶6–7, ECF No. 53. Three valley fills at Mine No. 2 drain into Outlet 001, which discharges into Road Fork. *Id.* at ¶¶3–5, 10.

Prior to Defendant's mining activities, Road Fork was without notable water quality issues. Collected in 1992 and 1993, pre-mining samples taken downstream from Outlet 001 and upstream from the confluence of Road Fork and Leatherwood Creek, showed conductivity levels ranging from 40 µS/cm to 73 µS/cm and sulfates ranging from 0.01 mg/l to 30 mg/l. Stipulation, ECF No. 53, Pls.' Ex. 73. In its 1994 Cumulative Hydrologic Impact Assessment (CHIA), WVDEP observed that "Road Fork does not appear heavily impacted by extensive past mining which has occurred in this area. This is indicated by low metals and sulfates that are less than 30 milligrams per liter." Pls.' Ex. 118 at PE1209–10.

Since Defendant's mining activities began, Road Fork water quality has notably suffered. Monitoring in 2010 through 2012 at that same discharge point revealed conductivity levels ranging

from 1,803 µS/cm to 5,700 µS/cm and sulfates ranging from 886 mg/L to 3,304 mg/L. In May 2014, the conductivity level of water discharged from Outlet 001 was 2,920 µS/cm and the sulfate level was 1,900 mg/L. Pls.' Exs. 2, 3. The following table provided by Dr. Palmer compiles the data collected at Fola Mine No. 4A discharges, including Broadtree Branch measurements as a comparison point for water chemistry characteristic of alkaline mine drainage. *See* Pls.' Ex. 38.

Location	**pH**	**Conductivity**	**Ca**	**Mg**	**Na**	**K**	**CL**	**SO_4**
Road Fork Outlet 001 (11/2006)	8.28	3290	358	310	12	n/a	2	419
Road Fork BASD-RFI (5/21/2011)	7.8	3200	385	382	11.9	10.7	18.5	1860
Road Fork BASD-RFI (5/21/2012)	7.98	2700	370	356	11.8	22.2	n/a	1860
Road Fork BASD-RFI (5/20/2013)	8.1	2530	330	320	12.7	19.1	n/a	1970
Road Fork Outlet 001 (Hansen 9/9/2014)	7.18	3370	290	300	10	17	ND	2100
Road Fork BASD-RFI (Hansen 5/19/2014)	7.18	3370	320	292	10.6	17.7	n/a	1620
Boardtree Branch	8	2367	241	260	12	21	11	1580

Table 4: Water quality data for Mine No. 2 discharges into Road Fork, including data from Boardtree Branch as a reference point.

Consistent with the observed decline in water quality, the aquatic community at Road Fork is biologically impaired. Between 2011 and 2014, Defendant reported WVSCI scores between 46 and 56 in Road Fork downstream of Outlet 001. Stipulation ¶17. These scores are well below the accepted EPA threshold marking impairment at WVSCI scores under 68. *See Elk Run*, 24 F.Supp.3d at 554–56 (discussing reliance on WVSCI scores lower than 68 as an agency-derived and federally approved marker of biological impairment); *see also*, *supra* note 6. On May 9,

2014, Dr. Christopher Swan conducted field sampling immediately downstream of Outlet 001. Based on that sampling effort, Dr. Swan identified a WVSCI score of 39.66 and a GLIMPSS score of 20.22. Pls.' Ex. 25. Both numbers clearly indicate impairment. *See* Pls.' Ex. 173 at JE0024; Tr. 2 at 89–90, 161, ECF No. 100; Tr. 3 at 36–37, June 3, 2015, ECF No. 106. Given these numbers, it is not at all surprising that WVDEP lists Road Fork and Leatherwood Creek as biologically impaired due to mining on its 2012 CWA 303(d) List. Joint Ex. 20; Tr. 2 at 149, ECF No. 100. Furthermore, WVDEP observed in its Elk River Watershed TMDL that ionic toxicity is a definite stressor at Road Fork. Joint Ex. 16 at 24; Tr. 2 at 149–50, ECF No. 100 ("In [Road Fork/Leatherwood Creek] . . ., the [stressor identification] process determined ionic toxicity to be a significant stressor. A strong presence of sulfates and other dissolved solids exists in those waters and in all other streams where ionic toxicity has been determined to be a significant biological stressor.").

In addition to conducting water sampling, multiple experts also provided assessments of habitat at Road Fork. Dr. Swan reported a 163 RBP habitat score in Road Fork. Pls.' Ex. 31. In contrast to Dr. Swan's optimal-range RBP score, Dr. Menzie also conducted or oversaw RBP assessments. The average RBP score taken by two Exponent staff members in October 2014 was 137.5, or suboptimal.[37] Def.'s Ex. 138. Drs. Swan and Menzie both reported the presence of metal hydroxides in some stretches of stream. Tr. 3 at 58, 67–69, ECF No. 106; Tr. 4 at 23, ECF No. 107.

[37] Whatever critiques one may have of RBP habitat assessment protocol, presumably all can agree that the experience and qualifications of the person performing the habitat assessment are extremely relevant to determining the reliability of that assessment. Though he deliberately deviated from the WVDEP protocol for conducting an RBP habitat assessment, the Court has no doubt that Dr. Menzie possesses the necessary experience and qualifications to render his assessment reliable. However, the Court heard nothing on the identity, experience, or qualifications of Exponent staff members that performed RBP habitat assessments at Dr. Menzie's request. Accordingly, the Court hesitates to rely on such assessments.

2. Fola Surface Mine No. 4A – Right Fork

Fola Surface Mine No. 4A is regulated under WV/NPDES Permit No. WV1013815 and West Virginia Surface Mining Permit S200502. Stipulation ¶¶23–24, 26, ECF No. 53. Plaintiffs specified three outlets in their complaint—Outlets 22, 23, and 27—each of which discharges into tributaries of Leatherwood Creek. *Id.* at ¶¶21–22.

Though there was some pre-Fola mining in the area, Right Fork was unimpaired and in fair shape prior to Defendant's mining operations. In its 2003 CHIA, the WVDEP observed that though some sub-watersheds of Right Fork had elevated Manganese and sulfates related to pre-Fola mining, the upper reaches of the watershed maintained low sulfates. Pls.' Ex. 89 at PE0574–75. Indicating high water quality, the report further noted that "all [monitoring] stations provide adequate habitat and contain populations of benthic macroinvertebrates. All the stations have high EPT indices." Pls.' Ex. 89 at PE 0577–78. The majority of pre-Fola water samples showed conductivity levels below the EPA threshold of 300 μS/cm, with some samples deviating upward as high as 1500 μS/cm. Pls.' Ex. 44.

Not surprisingly, the relatively good water quality at Right Fork contributed to unimpaired conditions. In 1997, WVDEP reported an excellent WVSCI score of 84 for Right Fork. Joint Ex. 23 at 68, Tr. 2 at 173, ECF No. 100. In 2000 and 2001, Fola's consultant collected a number of biological surveys from seventeen different sampling locations. Among thirty-three samples from those seventeen sites, only six returned WVSCI scores below 68. Stipulation ¶¶22, 29–31, Tr. 2 at 177, ECF No. 100.

Since Defendant's mining activities began, Right Fork water quality has notably suffered. Though jumps in conductivity had previously been rare, since 2001, conductivity levels in Right Fork have been almost entirely above 1,500 μS/cm, now with jumps up to and exceeding 2,500

μS/cm. Stipulation ¶33. Similarly, since 2001, sulfate levels are consistently above 600 mg/l; and sometimes as high as 1,200 mg/l. Stipulation ¶33.

Consistent with conditions in Right Fork itself, in 2011 and 2012, discharges from Outlets 022, 023, and 027 consistently ranged from 1,500 μS/cm to more than 3,000 μS/cm. Stipulation ¶32. In May and September 2014, conductivity from the three discharges ranged from 1820 to 2,958 μS/cm, with sulfate levels between 920 and 1,800 mg/l. Pls.' Ex. 2–5. The following table provided by Dr. Palmer compiles the data collected at Fola Mine No. 4A discharges, including Broadtree Branch measurements as a comparison point for water chemistry characteristic of alkaline mine drainage. *See* Pls.' Ex. 49.

	Location	**pH**	**Conductivity**	**Ca**	**Mg**	**Na**	**K**	**CL**	**SO_4**
Pre-Fola Mining	**FOLA – 6 (2001)**	7.15	461	34	25	8	3	3	120
	FOLA – 7 (2001)	7.35	367	34	75	2	3	1	110
Post-Fola Mining	**BASD3RLW (2012)**	8.38	1689	265	211	30	16	n/a	1150
	BASD1RLW (2012)	8.17	1538	202	156	31	14	n/a	942
	Outlet 022 (Hansen 2014)	7.9	1820	140	120	62	12	32	920
	Outlet 023 (Hansen 2014)	8.1	2720	280	260	100	16	ND	1800
	Outlet 027 (Hansen 2014)	7.12	2390	220	130	140	14	ND	1300
Reference	**Boardtree Branch**	8	2367	241	260	12	21	11	1580

Table 5: Water quality data for Mine No. 4A discharges into Right Fork, including data from Boardtree Branch as a reference point.

The noted decline in water quality has been accompanied by a decline in WVSCI scores. On May 9, 2014, Dr. Christopher Swan conducted field sampling downstream of Outlets 022, 023, and 027.[38] Based on that sampling effort, Dr. Swan identified a WVSCI score of 38.21 and a GLIMPSS score of 25.79. Pls.' Ex. 25; Tr. 2 at 184, ECF No. 100. Both numbers clearly indicate impairment and are a far cry from the pre-Fola mining scores. Given these numbers, it is not at all surprising that WVDEP lists Right Fork and Leatherwood Creek as biologically impaired due to mining on its 2012 CWA 303(d) List. Joint Ex. 20; Tr. 2 at 149, ECF No. 100. Furthermore, WVDEP observed in its Elk River Watershed TMDL that ionic toxicity levels in Right Fork are a definite stressor. Joint Ex. 16 at 24; Tr. 2 at 149–50, ECF No. 100 ("In [Right Fork/Leatherwood Creek] . . ., the [stressor identification] process determined ionic toxicity to be a significant stressor. A strong presence of sulfates and other dissolved solids exists in those waters and in all other streams where ionic toxicity has been determined to be a significant biological stressor.").

With respect to habitat, Dr. Swan calculated an RBP score of 172 at Right Fork, which is in the optimal range. Pls.' Ex. 32; Tr. 2 at 185, ECF No. 100. The average RBP score taken by two Exponent staff members in October 2014 was 128, or suboptimal. Def.'s Ex. 138.

3. Fola Surface Mine No. 6 – Cogar Hollow

Fola Surface Mine No. 6 is regulated under WV/NPDES Permit No. WV1018001 and West Virginia Surface Mining Permit S2011999. Stipulation ¶¶42–44, ECF No. 53. Three valley fills at Mine No. 6 drain into Outlets 013, 015, and 017, to be discharged to Cogar Hollow, a tributary of Leatherwood Creek. Stipulation ¶¶40–41; Tr. 2 at 187–88, ECF No. 100.

Before Defendant's mining activities began, Cogar Hollow enjoyed healthy water quality.

[38] As will be discussed in greater detail below, Dr. Swan's sampling area was not immediately downstream of Outlets 022, 023, and 027 and did not isolate the impacts of these three discharge outlets. Instead, Dr. Swan's sampling area accounted for as many as twelve additional outlets.

Measured at monitoring point S3-1A, conductivity levels were usually well below 300 μS/cm. Stipulation ¶45. Sulfate levels were similarly low. *Id.* As it did with other Leatherwood Tributaries, here WVDEP again noted that "all [monitoring] stations provide adequate habitat and contain populations of benthic macroinvertebrates. All the stations have high EPT indices." Pls.' Ex. 89 at PE 0577–78.

Since Defendant's mining activities began, Cogar Hollow water quality has diminished considerably, including extremely elevated conductivity and sulfate levels. Measurements taken since July 2012 consistently reveal conductivity levels ranging from 3,000 μS/cm to 5,000 μS/cm. Stipulation ¶47; Tr. 2 at 191–92, ECF No. 100. In May 2014, conductivity from the three discharges ranged from 2,910 μS/cm to 3,202 μS/cm and the sulfate level was 1,900 mg/l to 2,400 mg/l. Pls.' Ex. 2–3, 5. The following table provided by Dr. Palmer compiles the data collected at Fola Mine No. 4A discharges, including Broadtree Branch measurements as a comparison point for water chemistry characteristic of alkaline mine drainage. Pls.' Ex. 57.

Location	**pH**	**Conductivity**	**Ca**	**Mg**	**Na**	**K**	**CL**	**SO_4**
Mine No. 6 Outlets (July 2007)	6.03	3420	486	254	9	n/a	1	1912
Mine No. 6 Outlet 013 (June 5, 2012)	7.52	n/a	n/a	448	n/a	n/a	8.93	2786
Mine No. 6 Outlet 015 (June 5, 2012)	7.09	n/a	n/a	n/a	n/a	n/a	7.98	2018
Mine No. 6 Outlet 017 (May 23, 2012)	6.53	n/a	n/a	284	n/a	n/a	10.24	2133
Mine No. 6 Outlet 013 (Hansen 2014)	7.73	4200	360	400	63	20	ND	2700
Boardtree Branch	8	2367	241	260	12	21	11	1580

Table 6: Water quality data for Mine No. 6 discharges into Cogar Hollow, including data from Boardtree Branch as a reference point.

On May 9, 2014, Dr. Christopher Swan conducted field sampling immediately downstream of Outlets 013, 015 and 017. Based on that sampling effort, Dr. Swan identified a

WVSCI score of 41.81 and a GLIMPSS score of 20.03. Pls.' Ex. 25; Tr. 2 at 194, ECF No. 100. Both numbers clearly indicate impairment. Dr. Swan further conducted a RBP habitat assessment, scoring the stream at 145, or suboptimal. Pls.' Ex. 30; Tr. 3 at 44, ECF No. 106. Similarly, the average RBP score taken by two Exponent staff members in October 2014 was 138.5, again, suboptimal. Def.'s Ex. 138.

On the basis of the foregoing site-specific evidence, the Court **FINDS** by a preponderance of the evidence that Road Fork, Right Fork, and Cogar Hollow are biologically impaired.

B. Legal Standard

The Court must find a violation here if Defendant's discharges cause or materially contribute to a significant adverse impact to the chemical or biological components of aquatic ecosystems. W. Va. Code R. § 47-2-3.2.i. Through requiring that a discharge "cause or materially contribute" to biological impairment, West Virginia law imposes something less stringent than traditional but-for causation. Unfortunately, available state law does not elaborate on what precisely that standard means as applied.

This Court has previously ruled that "[i]t is readily conceivable that multiple pollutants or stream characteristics might simultaneously materially contribute to impairment; and [a plaintiff] need only provide evidence showing it is more probable than not that ionic pollution as measured by conductivity is among some collection of material contributors." *Fola (Stillhouse)*, 2015 WL 362643, at *8. This approach to "material contribution" is consistent with the Ninth Circuit's explanation that "material contribution" suggests that "more than one factor can be a substantial cause, and no single factor need be the sole causative element." *Frito-Lay, Inc. v. Local Union No.* 137, 623 F.2d 1354, 1363 (9th Cir. 1980); *accord Feather v. United Mine Works of America*, 903 F.2d 961, 967 (3rd Cir. 1990).

This Court's interpretation of the material contribution standard flows from the observation that, "as a matter of plain meaning and common sense, it is possible to identify a factor that is materially contributing to a given condition without conclusively eliminating contributions by additional factors in a dynamic system." *Fola (Stillhouse)*, 2015 WL 362643, at *9. Similarly, in the context of environmental litigation, it is reasonable to expect that biological communities may be simultaneously impaired by varied and multiple pollutants. Liability cannot be skirted by the mere presence of multiple stressors, lest we enable the simple nature of ecological systems to invariably frustrate the Clean Water Act.

While something less than traditional but-for causation is required, Defendant argues that, if the word "material" is to have any import, "material contribution" must amount to something more than a contributing factor test. Def.'s Post-Trial Brief, ECF No. 116 at 27 (citing *Artz v. Chicago, R.I. & P.R. Co.*, 38 Iowa 293, 296–97 (Iowa 1874)). *But see Coeur D'Alene Tribe v. Asarco Inc.*, No. CV91–0342NEJL, 2001 WL 34139603, at *4 (D. Idaho March 30, 2001) (considering liability for comingled pollutants under a contributing factor rest, the court observed that "[plaintiffs] have the burden of proving a release that results in commingled hazardous substances is a 'contributing factor' [more than a de minimum amount—to an extent that at least some of the injury would have occurred if only the Defendant's amount of release had occurred]."); *cf.* Sanne H. Knudsen, *The Long-Term Tort: In Search of a New Causation Framework for Natural Resource Damages*, 109 Nw. U. L. Rev. 475, 532 (Winter 2004) (explaining that the "[contributing factor] doctrine appears suited to handle problems of synergistic harms"). As argued by Defendant, "even if the conductivity [here] were fully treated, it is unclear what the resulting stream score would be given the other factors at issue." *Id.* at 28. This framing reflects a fundamental misunderstanding. Obtaining a WV/NPDES permit to

discharge into a given waterway does not transmute a permitee into a guarantor that the receiving stream has or will maintain a passing WVSCI score. More modestly, issuance of the permit requires that a permitee not cause or materially contribute to impairment; it does not require or guarantee nonimpairment generally.

Thus, the Court continues to require that Plaintiffs demonstrate that it is more probable than not that ionic pollution, measured as conductivity, is among some collection of material contributors to biological impairment. Conductivity levels in the streams at issue need not be the sole cause of observed biological impairment, but must be a substantial contributor. This standard does not require scientific certainty, but rather legal probability. *Fola (Stillhouse)*, 2015 WL 362643, at *17, *quoting Ferebee v. Co.*, 736 F.2d 1529, 1536 (D.C. Cir. 1984); *Bunting v. Secretary of Health & Human Services*, 931 F.2d 867, 873 (Fed. Cir. 1991).

C. Specific Causation Analysis

1. Defendant's Motion for a Directed Verdict regarding Fola Surface Mine No. 4A

At the close of Plaintiffs' case in chief, Defendant moved for a directed verdict as to claims involving Fola Surface Mine No. 4A, covered by WV/NPDES Permit No. WV1013815. Tr. 3 at 202, ECF No. 106. Defendant argues that it is entitled to a directed verdict regarding Fola Surface Mine No. 4A because Plaintiffs' compliance evidence does not isolate the effects of the discharges from Outlets 022, 023, and 027. *Id.* Outlets 022 and 023 at Mine No. 4A discharge into Right Fork, a tributary of Leatherwood Creek. Outlet 027 at Mine No. 4A discharges into Cannel Coal Hollow, a small tributary of Leatherwood Creek. Plaintiffs' consultant conducted biological sampling to determine a WVSCI score for Right Fork at a stream reach below the confluence of Cannel Coal Hollow. In addition to Outlets 022, 023, and 027, some twelve additional outlets discharge into Right Fork or Cannel Coal Hollow upstream of Plaintiffs' compliance sampling

location. As a result, Plaintiffs' sampling does not isolate the effect of discharges only from Outlets 022, 023, and 027. Tr. 2 at 199, ECF No. 100 (testimony of Dr. Palmer, agreeing that sampling did not isolate the effects of discharges only from Outlets 022, 023, and 027).

In answer to Defendant's oral motion, Plaintiffs maintained that, under a material contribution standard, they presented sufficient evidence to show that high conductivity discharges from the three outlets are materially contributing to the impairment observed downstream in Right Fork. Tr. 3 at 203, ECF No. 106. Plaintiffs report that the combined flow from Outlets 022, 023, and 027 is approximately 400 to 420 gallons per minute, or roughly 576,000 to 604,800 gallons per day. In April 2012, WVDEP reported that the total flow at the mouth of Right Fork is approximately 7.49 cubic feet per second, or roughly 4.5 million gallons per day. Def.'s Ex. 198 at FOLA#4A000986. Relying on those flow estimates, Plaintiffs calculate that Outlets 022, 023, and 027 contribute one-eighth or more of the total Right Fork flow. As reasoned by Plaintiffs, though their compliance sampling did not isolate the effects of Outlets 022, 023, and 027, a one-eighth contribution to the total flow is sufficient to meet the material contribution standard.

The Court disagrees. While "material contribution" does not require evidence that a single stressor independently causes impairment, it must require something more than what Plaintiffs have demonstrated here. First, the flow evidence offered by Plaintiffs is thin at best. It does not account for seasonal variations or for the variability of inputs from rainfall. Moreover, the Court heard no expert testimony to suggest that the some 600,000 gallons per day coming from the three outlets at issue materially contributes to impairment observed downstream after the introduction of nearly four million additional gallons from sources unknown, but including at least twelve other discharge outlets—each of which is a potential source of ionic pollution. Without supporting scientific testimony, the Court cannot connect the discharges at issue to the area where

Plaintiffs performed compliance sampling. Doing so would be little more than non-scientific speculation. Thus, the Court cannot conclude that Plaintiffs have demonstrated that Outlets 022, 023, and 027 are materially contributing to downstream impairment.

Accordingly, the Court **FINDS** that Plaintiffs have not met their burden of proving that discharges from Outlets 022, 023, and 027 cause or materially contribute to biological impairment of Right Fork, and Defendant is therefore entitled to judgment as a matter of law regarding alleged violations of WV/NPDES Permit No. WV1013815, governing Fola Surface Mine No. 4A.

2. Liability with respect to Fola Surface Mine No. 2 and No. 6.

Having found that both Road Fork and Cogar Hollow are biologically impaired, the Court finally turns to the question of whether Plaintiffs' have met their burden in establishing that high conductivity discharges from Defendant's Mine No. 2 and Mine No. 6 cause or materially contribute to impairment.

Plaintiffs have sufficiently demonstrated that the discharges in question share the characteristic ionic composition identified and analyzed by the EPA in its Benchmark. Dr. Palmer testified that the ionic composition of the discharges matches the characteristic ionic composition associated with alkaline mine drainage in the region (e.g., sulfates, bicarbonates, calcium, and magnesium). Stipulation ¶¶18, 36, 38, 48; Pls.' Exs. 38, 49, 57; Tr. 2 at 93, 156–57, 181–84, and 192–93, ECF No. 100. To demonstrate the similarity, Dr. Palmer compared water quality measurements at each discharge outlet to the composition of reconstituted water from Boardtree Branch—recognized as representing the particular ionic composition of regional alkaline mine drainage. *See supra* Tables 4, 5, and 6; Kunz et al., *Use of Reconstituted Waters to Evaluate Effect of Elevated Major Ions Associated with Mountaintop Coal Mining on Freshwater Invertebrates*, 32 Envtl. Toxicology & Chemistry 2826 (2013), Joint Ex. 10. Testing the

characteristically composed water, Kunz et al., reported adverse effects to the mayfly *Isonychia* at conductivity levels of 1090 μS/cm—well below the conductivity levels at issue here.

Additionally, Plaintiffs have unequivocally demonstrated that the conductivity levels at issue here are sufficiently high to cause the observed impairment. Discharges from Fola Surface Mine No. 2 and No. 6 consistently and grossly exceed the threshold identified by EPA at which it becomes more likely than not that a stream will suffer biological impairment. *See Mancuso v. Consolidated Edison Co. of New York, Inc.*, 56 F.Supp. 391, 403 (S.D.N.Y. 1999) ("A fundamental tenet of toxicology is that the 'dose makes the poison' and that all chemical agents, including water, are harmful if consumed in large quantities, while even the most toxic substances are harmless in minute quantities.") aff'd in part, vacated in part, 216 F.3d 1072 (2nd Cir. 2000). Here, Defendant's discharges are consistently in the range of 2,000 μS/cm to 3,000 μS/cm, and are sometimes as high as 4,000 μS/cm to 5,000 μS/cm. Pls.' Ex. 37, 45, 55; Tr. 2 at 155–56, 178, 191, ECF No. 100. Those conductivity levels are several times the threshold identified by an expert federal agency as well as thresholds independently researched and reported in peer-reviewed scientific journals.[39]

Sampling at Road Fork and Cogar Hollow revealed changes in the macroinvertebrate community that are consistent with impairment caused by high conductivity levels. Multiple peer-reviewed scientific articles report finding that high conductivity associated with alkaline mine drainage leads to the extirpation of mayflies. Pls.' Ex. 173 at PE1536, JE0010, PE1832; Tr. 2 at 105–07, 137, ECF No. 100. Here, consistent with those reported findings, sampling revealed the complete extirpation of mayflies. Tr. 3 at 162–64, ECF No. 106 (noting that the complete

[39] Though not required for a finding of violation, the Court notes that the evidence presented is arguably sufficient to establish that though conductivity may not be the sole driver of impairment here, conductivity levels such as these are likely capable of causing impairment even standing alone.

extirpation of mayflies—organisms known to be particularly sensitive to conductivity—is strong evidence that conductivity is driving impairment); Joint Ex. 1 ("The clear patterns linking high conductivity to a loss of mayfly taxa has ecosystem-scale importance since mayflies often account for 25-50 percent of total macroinvertebrate abundance in the least-disturbed Central Appalachian streams.").

On the basis of the evidence presented, the Court **FINDS** that Plaintiffs have demonstrated the following by a preponderance of the evidence: (1) Road Fork and Cogar Hollow are biologically impaired, as measured by WVSCI scores below the federally approved threshold score of 68; (2) the discharges complained of share the same characteristic composition presently known in the scientific community to cause or materially contribute to impairment in central Appalachian streams; (3) the conductivity levels observed at Defendant's discharges are far in excess of the thresholds identified by the EPA and the available scientific literature known to cause stress to aquatic communities; (4) Defendant's mining operations are the only land use that could have caused impairment; and (5) changes in the biological community particularly show the loss of conductivity-intolerant organisms. Thus, the Court **FINDS** that high conductivity discharges from Fola Surface Mine No. 2 and No. 6 are causing or materially contributing to the biological impairment of Road Fork and Cogar Hollow in violation of Defendant's current WV/NPDES permits.

These findings are consistent with WVDEP's determination that ionic toxicity is a "significant stressor" in both Road Fork and Right Fork. Joint Ex. 16 at JE0578 ("A strong presence of sulfates and other dissolved solids exists in those waters and in all other streams where ionic toxicity has been determined to be a significant biological stressor."

Moreover, these findings are consistent with the testimony of Dr. Menzie, Defendant's only expert testifying on specific causation. Dr. Menzie offered consistent and repeated testimony that conductivity is one among a small collection of substantial causes of impairment. Tr. 3 at 240, ECF No. 106 ("[Conductivity is] going to play some role."); Tr. 4 at 104, ECF No. 107 ("I include conductivity as a stressor along with that"); Tr. 4 at 76, ECF No. 107 (listing precipitates, temperature, and conductivity as the drivers of impairment at the Leatherwood tributaries); Tr. 4 at 98-100, ECF No. 107 (explaining that conductivity is one of the factors causing impairment, but adding that he did not think it possible to allocate contribution more precisely). Though inconsistent with all other scientific evidence addressing thresholds, Dr. Menzie further maintained the opinion that conductivity levels in the range of 1,000 μS/cm to 3,000 μS/cm are capable of independently influencing WVSCI scores. Tr. 4 at 100, ECF No. 107. Even assuming Dr. Menzie's idiomatic threshold is correct despite all evidence to the contrary, this case concerns discharges with conductivity levels in that range and well above. Thus, WVDEP and all testifying experts in this case share the opinion that conductivity levels are causing or materially contributing to biological impairment.

3. Defendant's Rebuttal Arguments

Defendant advances a variety of arguments designed to defeat Plaintiffs' theory of specific causation. These arguments question the reliability and credibility of Plaintiffs' experts, the adequacy of Plaintiff's causation evidence, and the possibility of alternative drivers of impairment. Notwithstanding the fact that experts for both parties agree that conductivity levels are materially contributing to the observed biological impairment of Road Fork and Cogar Hollow, counsel continues to piecemeal attack Plaintiffs' evidence on specific causation.

As stated by Defendant, with regard to specific causation, "[d]etermining which factors are

material is the stuff of scientific investigation, not assumption." ECF No. 116 at 8. Here, Plaintiffs' experts, Defendant's expert, and the WVDEP all agree: high conductivity is materially contributing to biological impairment at Road Fork and Cogar Hollow. The Court will not displace these expert opinions based on independent scientific investigation in favor of lay assumptions. Accordingly, the Court declines to delve further into counsel's evidence-based arguments; leaving only Defendant's legal argument regarding the adequacy of notice of potential liability.

Contrary to Defendant's assertion, in 2008, when Defendant's permits were last renewed, Defendant had notice of potential liability should it cause or contribute to biological impairment. Defendant's permits incorporate § 47-30-5.1.f., an unambiguous permit condition imposing liability should a permitee violate West Virginia's narrative water quality standards. *See OVEC v. Elk Run*, 2014 WL 29562, at *10. At the earliest, permittees have had notice of such potential liability since the predecessor to § 47-30-5.1.f. first became effective in 1985, or at the latest, upon original issuance or renewal of a particular permit. *Id.*

Not only did Defendant therefore have notice of potential liability provided in the plain language of its permits and each subsequent reissuance, but at the time of reissuance, Defendant further had the benefit of scientific literature demonstrating the harmful effects of high conductivity. As reviewed above, in a 2003 Programmatic Environmental Impact Statement on Mountaintop Mining/Valley Fills in Appalachia, the EPA reported a known association between increases in conductivity and coincident downstream biological impairment.[40] The EPA's

[40] Commenters on the draft EIS included several representatives for the West Virginia Coal Association, as well as representatives on behalf of the Kentucky Coal Association and Ohio Coal Association. EPA, Public Comment Compendium: Mountaintop Mining/Valley Fills in Appalachia Final Programmatic Environmental Impact Statement, Vol. 1 (Oct. 2005) (available at: http://www.epa.gov/region3/mtntop/documents.htm#cd).

Environmental Impact Statement was followed by additional publications in the peer-reviewed scientific literature well before Defendant's 2008 permit reissuance. *See supra* Section II.C. Accordingly, the Court concludes that Defendant had the benefit of adequate notice.

IV. CONCLUSION

In sum, the Court **FINDS** that Plaintiffs have established, by a preponderance of the evidence, that the Defendant has committed at least one violation of its permits for Mine No. 2 and Mine No. 6 by discharging into Road Fork and Cogar Hollow high levels of ionic pollution, which have caused or materially contributed to a significant adverse impact to the chemical and biological components of the applicable streams' aquatic ecosystem, in violation of the narrative water quality standards that are incorporated into those permits.

The Court **DIRECTS** the Clerk to send a copy of this written Opinion and Order to counsel of record and any unrepresented parties.

ENTER: August 12, 2015

ROBERT C. CHAMBERS, CHIEF JUDGE

Given the prevalence of such commenters, it is difficult to imagine that Fola could have remained uninformed of the EPA's draft EIS. To the contrary, it more likely suggests regional, industry-wide awareness.