

United States Court of Appeals

FOR THE DISTRICT OF COLUMBIA CIRCUIT

Argued October 15, 2004

Decided December 10, 2004

No. 96-1062

EDISON ELECTRIC INSTITUTE, ET AL.,
PETITIONERS

v.

ENVIRONMENTAL PROTECTION AGENCY, ET AL.,
RESPONDENTS

AMERICAN PETROLEUM INSTITUTE,
INTERVENOR

Consolidated with
Nos. 96-1124, 03-1087, 03-1091, 03-1094

On Petitions for Review of an Order of the
Environmental Protection Agency

John C. Hall argued the cause for petitioners. With him on the briefs were *James N. Christman*, *Andrea W. Wortzel*, *Alexandra D. Dunn*, *David E. Evans*, *Stewart T. Leeth*, and *Richard H. Sedgley*. *Kristy A.N. Bulleit* entered an appearance.

Fredric P. Andes argued the cause and filed the briefs for intervenor.

Christina B. Parascandola and *David S. Gualtieri*, Attorneys, U.S. Department of Justice, argued the cause and filed the brief for respondents. *Daniel R. Dertke*, Attorney, U.S. Department of Justice, entered an appearance.

Before: EDWARDS and RANDOLPH, *Circuit Judges*, and WILLIAMS, *Senior Circuit Judge*.

Opinion for the Court filed by *Circuit Judge* RANDOLPH.

RANDOLPH, *Circuit Judge*: Edison Electric Institute and organizations representing corporate and municipal dischargers brought these consolidated petitions for review, claiming that certain of EPA's "whole effluent toxicity" or "WET" test methods were invalid. The tests are set forth in rules promulgated pursuant to the Clean Water Act, 33 U.S.C. § 1251 *et seq.* (the "Act"). The Act prohibits the discharge of pollutants except in compliance with individual permits issued by EPA or the states. States prescribe their own water quality criteria, which EPA reviews for conformity with the Act. Water quality standards typically consist of two complementary parts: numerical limits on the allowable concentration of particular pollutants in ambient water (*e.g.*, "no more mercury than 5 parts per billion"), and a descriptive, "narrative" criterion regarding the entire effluent (*e.g.*, "no toxic pollutants in toxic amounts"). See 33 U.S.C. § 1251(a)(3). WET tests are used to measure compliance with standards of the latter type.

While the numerical restrictions comprise the backbone of the permitting system, EPA has found that, standing alone, these limits are not sufficient. Effluents may contain many different pollutants. Even if no single pollutant were present in a harmful amount, the mix of different pollutants still might have negative effects upon aquatic organisms. In light of the myriad potential interactions among various pollutants,

traditional instrumental tests are ill-suited to making the determination. Instead, laboratories expose aquatic organisms to samples of the effluent, at various concentrations, and measure the extent to which the organisms are adversely affected. If, in the laboratory, the effluent is harmful to the test organisms at a certain concentration, then it is presumed also to be harmful to aquatic life in the stream—*i.e.*, to be toxic—at that concentration.

This approach has an appealing simplicity, but the use of living specimens introduces a significant potential for variability between and within tests. In designing and refining the WET test methods, EPA sought to minimize the effect of organic idiosyncrasy by taking experimental and statistical precautions. The crux of petitioners' complaint is that EPA has not gone far enough. We disagree, and therefore deny the petitions for review.

I.

These WET test methods were first implemented in 1995. 60 Fed. Reg. 53,529 (Oct. 16, 1995). Petitioners brought an action challenging them, as a result of which the WET tests were modified pursuant to a settlement of the action, after which EPA repromulgated them in 2002. 67 Fed. Reg. 69,952 (Nov. 19, 2002) ("Final Rule").¹ It is this most recent version of the

¹ Petitioners object to four of the ten test procedures described in the 2002 Final Rule: the Fathead Minnow Larval Growth Test Method 1000.0, the Fathead Minnow Embryo-larval Teratogenicity Test Method 1001.0, *Ceriodaphnia dubia* (water flea) Reproduction Test Method 1002.0, and Green Alga Growth Test Method 1003.0. See 67 Fed. Reg. at 69,972. Each of these four tests measures chronic toxicity, which is defined in relation to test organisms' growth and reproduction, as opposed to acute toxicity, which is based on mortality

tests that we now review.

A.

Petitioners' primary concern is that EPA did not adhere to its usual criteria and procedures for ensuring the scientific validity of the test methods.² These criteria include accuracy, precision, practical applicability, establishment of detection limits, and the minimization of external interference. *See EPA, Availability, Adequacy, and Comparability of Testing Procedures for the Analysis of Pollutants Established Under Section 304(h) of the Federal Water Pollution Control Act 3-2 to 3-5 (Sept. 1988) ("Report to Congress").*³ While EPA concedes that its WET tests do not incorporate every one of

rates. *Id.* at 69,953.

² Petitioners suggest, without supporting authority, that because the test results will be used as evidence in enforcement proceedings, EPA's rulemaking had to comply with the standard for scientific evidence articulated in FED. R. EVID. 702, as interpreted in *Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 579 (1993). Evidentiary rules govern the admissibility of evidence at trial, not the establishment of the processes whereby such evidence will be created. *See* FED. R. EVID. 101 ("These rules govern proceedings in the courts of the United States . . ."). Of course, insofar as some of EPA's own criteria mirror the *Daubert* standard, EPA may not ignore or contradict them without explanation.

³ This Report was an internal study on various testing methods, undertaken at Congress's express behest. Pub. L. No. 100-4, § 518(a), 101 Stat. 7, 86-87 (1987). The Report itself nowhere contemplates being anything but a "study." It is not strictly binding upon EPA and any deviation from the Report is not *per se* arbitrary and capricious. *Cf. Report to Congress* at 3-2 ("In most cases, no single [test] method will contain all of the desirable characteristics.").

these factors, the real question is whether EPA adequately accounted for any departures. We find that it did.

EPA explained at length, both in its response to public comments and in the Final Rule, that there are two major distinctions between WET tests and most other test methods approved for assessing permit compliance under the Act. First, while most tests rely on instrumentation to conduct chemical-specific numerical measurements, WET testing is biological, using live organisms that cannot be, for example, calibrated. Second, unlike properties such as chemical concentration, toxicity is both measured and *defined* by the WET tests (*i.e.*, it is a “method-defined analyte”). These are meaningful differences, which serve to limit the usefulness of petitioners’ analogies between WET testing and chemical-specific instrumental methods.

EPA admits that accuracy, in its technical rather than colloquial sense, is inapplicable to WET testing, but it does not follow that the tests are therefore “inaccurate.” Accuracy is a composite of two distinct characteristics: “precision” and “bias.” The former measures the variation among the results of multiple tests of the same sample; the latter describes any systemic and persistent deviation of the average value of a test method from an accepted “true value.” Final Rule, 67 Fed. Reg. at 69,965. While precision can be, and has been, evaluated for WET methods, “bias” cannot be because it relies on comparisons with an independent, objective, “true value.” When measuring chemical concentration, for example, it is a simple matter for a laboratory to combine pure water with a given toxicant in a certain ratio, and then assess the ability of instruments correctly to ascertain this known concentration. For a method-defined analyte such as toxicity, however, there is no such thing as a “true value” independent of the WET tests themselves. This does not mean that the tests are inherently

unreliable, but rather that their scientific validity must be assessed through other means. This is consistent with EPA's treatment of other method-definite analytes. *See generally* 40 C.F.R. pt. 136.

While conceding the inapplicability of bias, EPA stated in the rulemaking that its WET test methods satisfy precision. 67 Fed. Reg. at 69,965. Petitioners argue that this conclusion is unsupported. The record contains extensive raw data, from the main EPA Interlaboratory Study and other privately commissioned studies, regarding the variability of WET toxicity measurements. *See, e.g., EPA, Final Report: Interlaboratory Variability Study of EPA Short-term Chronic and Acute Whole Effluent Toxicity Test Methods* (Sept. 2001) ("Interlaboratory Study"). From essentially the same data, petitioners draw quite different statistical conclusions than EPA.

Petitioners' analysis of this data does not convince us that EPA's action was "arbitrary, capricious, an abuse of discretion, or otherwise not in accordance with law." 5 U.S.C. § 706(2)(A). And this is not just because of the deference we give to EPA when it evaluates "scientific data within its technical expertise." *City of Waukesha v. EPA*, 320 F.3d 228, 247 (D.C. Cir. 2003) (quoting *Huls Am., Inc. v. Browner*, 83 F.3d 445, 452 (D.C. Cir. 1996)); *Appalachian Power Co. v. EPA*, 135 F.3d 791, 801-02 (D.C. Cir. 1998). It is also because there are several errors in petitioners' methodology. One is petitioners' choice of units of measurement. According to EPA procedure, WET test results are recorded as percentages, representing how much dilution, if any, of an effluent sample is required for a certain effect to occur (*e.g.*, for the "No Observable Effect Concentration" datapoints, the percentage represents the level of dilution at which the mixture ceases to affect the organisms). Effluent that must be diluted to a 25% concentration before it ceases to cause demonstrable harm is

more toxic than effluent that need only be diluted to 50%. In order to simplify the expression and application of these test results, EPA devised a scale of chronic toxicity units (“TU_c”), equal to 100 divided by the measured percentage value, such that the 25% sample above would translate to 4 TU_c, while the 50% sample would be 2 TU_c. Thus, the higher an effluent’s TU_c rating, the more toxic the effluent. Petitioners make the mistake of assuming that relying on this invented scale in performing statistical analysis will yield valid conclusions about the distribution of the original data.⁴ This error lies at the heart of petitioners’ claims of extreme variability in the results of WET testing. EPA, on the other hand, finds that the data support the conclusion that these WET test methods exhibit a degree of precision compatible with numerous chemical-specific tests already in use. We credit EPA’s conclusions on this point.

Another of petitioners’ central contentions is that the WET test methods produce an unacceptably high number of false positives. EPA’s test design had contemplated a positive error rate of no more than five percent, and as low as one percent in certain instances; this understanding was reflected in

⁴ The preferred metric for assessing precision is the coefficient of variation (CV), which measures the extent to which multiple measurements tend to depart from their average value. The greater the CV, the less precise the measurement. By computing the CV using toxicity units (TU_cs) rather than the percentages originally recorded by EPA, petitioners arrive at a grossly inflated result. For example, analyzing reference toxicant data, *Interlaboratory Study* at 81-82 tbl. 9.8, EPA’s approach yields a CV of approximately 0.43—well within the range of EPA’s other approved tests, *Memorandum from Marion Kelly, EPA Engineering and Analysis Division 1* (Oct. 16, 2002) (CVs of approved chemical methods range from 0.03 to 0.64, and CVs of organic methods from 0.12 to 1.04). Petitioners’ approach, however, using the distorting TU_c scale, results in a CV of 1.47—more than triple the correct value.

the 1998 Settlement Agreement. Petitioners allege false positive rates between 12.5% and 56%, Reply Brief at 27, while EPA, again analyzing the same data, finds an overall false positive rate of 1.3%, with no individual test's rate exceeding 5%. See Final Rule, 67 Fed. Reg. at 69,968. The discrepancy stems from the parties' differing definitions of the term "false positive." EPA defines a false positive result as one indicating toxicity in a blank sample. *Interlaboratory Study* at 66. Such results occur quite infrequently. Petitioners' definition, however, is far more expansive, encompassing all results that exhibit toxicity greater than the median toxicity for a given sample. Reply Brief at 25 n.29. Their concern is that some discharge permits may specify an acceptable nonzero level of toxicity, which the effluent may not exceed, and that the WET tests have the potential to produce arbitrary permit violations. For example, if a permittee were subject to a toxicity limit of 3 TU_c, and a WET test of its effluent would yield a 2 TU_c result most of the time, but up to 4 TU_c some of the time, the latter outcome would constitute a permit violation and potentially trigger an EPA enforcement action.

This is certainly a problem for which EPA's system must account. It is not, however, a problem of false positives. What petitioners describe relates to precision, which we already have discussed. Multiple measurements will exhibit some degree of variation, yielding an error band that extends above and below some intermediate value. This is the case with chemical-specific instrumental tests and, indeed, with virtually every water quality test EPA uses. See 40 C.F.R. pt. 136. Furthermore, petitioners neglect to mention that just as some permittees who "should be" in compliance may be deemed violators, other permittees who "should be" violators may be deemed in compliance. That is the nature of any distribution: No matter how narrow the error band, or how precise the test, there always will be some measurements on the high end of the range, and some on the

low. The real question is whether this variation is excessive, and EPA has demonstrated that it is not. EPA also offered an additional safeguard by designing the tests to give permittees the benefit of the doubt, limiting false positive rates to at most 5%, while allowing false negative rates up to 20%. EPA, *Understanding and Accounting for Method Variability in Whole Effluent Toxicity Applications Under the National Pollutant Discharge Elimination System* 5-6 to 5-7 (June 2000).

It is worth pausing here before we examine petitioners' other attacks on the WET test methods. There is an important distinction between the validity of a test method and the validity of a particular result from the test when it is used to determine compliance with permit conditions. Even by EPA's calculations, WET tests will be wrong some of the time, which is why EPA warned against using a single test result to institute an action for a civil penalty. See 67 Fed. Reg. at 69,968. Nothing we have written thus far, and nothing we write in the balance of this opinion forecloses consideration of the validity of a particular test result in an enforcement action. See 33 U.S.C. § 1369(b)(2). That issue is not before us. The case involves only the validity of the WET test methods.⁵

⁵ One page of petitioners' opening brief contains what purports to be a constitutional argument—that if a particular WET test indicates toxicity, this will constitute an irrebuttable presumption of petitioners' guilt in violation of the Due Process Clause. As we stated in the text, we are concerned here only with test methodology, not results of particular tests in the field. Our decision does not endorse the validity of any test result in the future, nor does it foreclose a defense that the result is wrong. Those issues are simply not presented in this judicial review of rulemaking. Furthermore, when the Supreme Court has recognized the constitutional dimensions of presumptions, it has done so solely with regard to statutory classifications, which tended to have strong equal protection components as well. See, e.g., *Weinberger v. Salfi*, 422 U.S. 749 (1975) (Social Security eligibility classifications

Petitioners' next objection is to EPA's failure to establish detection limits for WET test methods. The public commenters raised this point and EPA explicitly addressed it in promulgating its Final Rule. 67 Fed. Reg. at 69,968. Detection limits are applicable only to tests that rely on instrumental measurements; they represent the sensitivity thresholds of the technology, below which measurements become unreliable or impossible. Because WET testing is a biological and experimental, rather than an instrumental, method, "detection limit concepts are not applicable." *Id.*; see also *Report to Congress* at 3-11. The ratified test methods, however, entail a built-in mechanism that serves the same basic purpose as detection limits in instrumental tests—to reduce the likelihood that random "noise" will result in a false positive result. A single WET test involves exposing multiple batches of organisms to the effluent at various concentrations, as well as to a "control" sample of pure water, and then aggregating the effects on each batch. Statistical analysis then is used to ensure that any observed differences between the organisms exposed to a given effluent concentration and those exposed to the control blanks most likely are not attributable to randomness—that they are statistically significant. See Final Rule, 67 Fed. Reg. at 69,957-58. This safeguard addresses petitioners' concerns. EPA, in short, has offered a reasoned and thorough explanation of its decision on

for spouses and stepchildren); *Vlandis v. Kline*, 412 U.S. 441 (1973) (state residency classifications for college tuition); see also JOHN E. NOWAK & RONALD D. ROTUNDA, CONSTITUTIONAL LAW § 13.6 (5th ed. 1995). There is no such classification here. To the extent petitioners' complaint is that in some future enforcement proceeding, they will not be able to attack the WET test methodology (if we rule in EPA's favor in this case), they are not speaking of an irrebuttable presumption at all. This case is their chance to rebut the so-called "presumption." Their inability to do so in some future proceeding is simply a consequence of the judicial review provision in 33 U.S.C. § 1369(b)(2).

this subject. The law requires no more. *See, e.g., Int'l Fabricare Inst. v. EPA*, 972 F.2d 384, 389 (D.C. Cir. 1992).

Petitioners also assert that EPA failed to demonstrate the availability and applicability of WET testing—that is, the ability of laboratories across the nation to conduct WET testing properly and consistently. One of the main purposes of the Interlaboratory Study was to ensure that a wide range of laboratories could implement the prescribed test methods without introducing an undue degree of variability or error. More than 90% of laboratories were able to complete the ratified tests in accordance with all mandatory procedures, with success rates reaching 100% for several tests. Final Rule, 67 Fed. Reg. at 69,955. When EPA was unable to find enough available laboratories for a trial of certain WET test methods, it withdrew those methods from 40 C.F.R. pt. 136. *Id.* Although the Interlaboratory Study clearly supports the availability and applicability of the challenged tests, petitioners think that procedural defects invalidate it. The claim is that because laboratories chosen for the test knew in advance that they would be participating, EPA violated its own guidelines, which required the study to be “blind.” *Interlaboratory Study* at A-21. This misapprehends the nature of blind testing. EPA called for “blind samples,” *id.* (emphasis added), and that is what the laboratories received—samples with no indication about which were the control “blanks” and which were the reference toxicants, Proposed Rule, 66 Fed. Reg. 49,794, 49,806 (Sept. 28, 2001). Petitioners also allege that EPA improperly ignored the results of the peer review process. But EPA published an extensive point-by-point response to peer comments and acknowledged the peer-review process in its revisions to the Final Rule, 67 Fed. Reg. at 69,954. The Interlaboratory Study thus complied with the appropriate procedures and established the ratified tests’ availability and applicability.

Another important test characteristic is “representativeness,” that is, the ability of test results to predict instream effects accurately. Petitioners claim that EPA failed to establish the presence of such correlations for several of the WET tests, particularly with regard to Western state waters, which differ chemically from their Eastern counterparts. EPA responds by pointing to the results of numerous studies on this subject conducted throughout the 1990s. These studies support the representativeness of the WET test methods in general, and several demonstrate representativeness with regard to particular Western waters. See, e.g., EPA, *A Review of Single Species Toxicity Tests: Are the Tests Reliable Predictors of Aquatic Ecosystem Community Responses?* 47-50 (July 1999). It is unrealistic in the extreme to require correlation studies on every stream in the nation. EPA took the sensible approach of relying on sampling techniques to draw general conclusions, while leaving some implementation details to local entities. See *Am. Iron & Steel Inst. v. EPA*, 115 F.3d 979, 1005 (D.C. Cir. 1997). Pursuant to the Clean Water Act’s National Pollutant Discharge Elimination System, 33 U.S.C. § 1342(a), states retain discretion, subject to EPA guidance and recommendations, to set their toxicity thresholds in order to compensate for local conditions at the permitting stage. See 40 C.F.R. § 122.44(d)(1)(iii). In light of this discretionary, rather than mandatory, nature of state implementation of standards and thresholds, we also are unpersuaded by petitioners’ assertion that the WET program amounts to an illegal federal water quality standard.

The role of state permitting authorities also should allay the concern, which petitioners express, that the correlation between laboratory toxicity and instream impacts grows weaker at lower levels of toxicity. Before implementing a test method, EPA must establish that the measured characteristic bears a rational relationship to real-world conditions; the available

studies reasonably support such a conclusion with regard to chronic toxicity. EPA, *Technical Support Document for Water Quality-Based Toxics Control* 8 (Mar. 1991) (finding likelihood that data may be explained by randomness, rather than actual correlation, to be 0.1%). Petitioners are worried that they might be subject to excessive restrictions; such limits, however, would be imposed by local authorities, and are not part of the rulemaking under review in this case. The WET test methods offer only a means of measuring compliance with those limits—individual dischargers remain free to challenge their permits, on a case-by-case basis, if they believe that local authorities are regulating at a level that poses only a minimal risk to aquatic life. See 40 C.F.R. §§ 124.19, 124.21.

The ratified WET tests are not without their flaws. But perfection is not the standard against which we judge agency action. *WorldCom, Inc. v. FCC*, 238 F.3d 449, 461 (D.C. Cir. 2001); *Northwest Airlines, Inc. v. U.S. Dep't of Transp.*, 15 F.3d 1112, 1119 (D.C. Cir. 1994). EPA's decision was informed by years of scientific studies, negotiation, and public notice-and-comment, and it represents the agency's expert judgment regarding the implementation of the aims of the Clean Water Act. Petitioners have not demonstrated that EPA ignored relevant record evidence, contradicted its own policies without explanation, or otherwise acted arbitrarily and capriciously. See *Motor Vehicle Mfrs. Ass'n v. State Farm Mut. Auto Ins. Co.*, 463 U.S. 29, 41-42 (1983); *Prof'l Pilots Fed'n v. FAA*, 118 F.3d 758, 771 (D.C. Cir. 1997); *Natural Res. Def. Council, Inc. v. EPA*, 822 F.2d 104, 111 (D.C. Cir. 1987).

II.

American Petroleum Institute ("API") seeks to intervene for the purpose of challenging EPA's *failure* to ratify for use in the Pacific Ocean three WET test methods that measure acute

toxicity. “An intervening party may join issue only on a matter that has been brought before the court by another party.” *Ill. Bell Tel. Co. v. FCC*, 911 F.2d 776, 786 (D.C. Cir. 1990), *citing* *Vinson v. Wash. Gas Light Co.*, 321 U.S. 489, 498 (1944). The issue presented by API overlaps with the issues petitioners raise only insofar as both involve whole effluent toxicity. The bare assertion that API “agree[s]” with petitioners’ claims, Reply Brief at 37, does little to cure this defect. The procedural device of intervention does not contemplate so broad a compass. We will not consider API’s arguments.

III.

For the reasons set forth above, having considered and rejected petitioners’ other arguments, we deny the petitions for review.