



MICHIGAN JOURNAL
of Environmental & Administrative Law

Spring
2013

Vol. 2
No. 2

ARTICLE

Making Method Visible:
Improving the Quality of
Science-Based Regulation

*Pasky Pascual, Wendy Wagner
& Elizabeth Fisher*

MAKING METHOD VISIBLE: IMPROVING THE QUALITY OF SCIENCE-BASED REGULATION

Pasky Pascual, Wendy Wagner** & Elizabeth Fisher****

Scientific inferences are theories about how the world works that scientists formulate based on their observations. One of the most difficult issues at the intersection of law and science is to determine whether the weight of evidence supports one scientific inference versus other competing interpretations of the observations. In administrative law, this difficulty is exacerbated by the behavior of both the courts and regulatory agencies. Agencies seldom achieve the requisite visibility that explains the analytical methods they use to reach their scientific inferences. Courts—because they appreciate neither the variety of inferential methods nor their epistemic foundations—do not demand this level of visibility from the agencies.

We argue that much progress can be made toward visible, coherent, science-based regulations if courts ask two deceptively simple questions: (1) have the agency's inferential methods been identified? and (2) does the agency explain how its methods are appropriate to the information on hand and how the methods support the agency's inferences?

INTRODUCTION	430
I. METHODS AND WHY THEY MATTER.....	436
A. <i>What Are the Methods of Scientific Inference?</i>	436
B. <i>How Was FDA Constrained to Rely on the Method of P-values as Its Sole Method of Making Inferences?</i>	440
C. <i>Invisible Methods in Practice</i>	445
1. Incorporating Methods into Pesticide Registration and Species Protection	445
2. Incorporating Methods into How Categories of Regulated Industries Are Defined.....	448
3. Incorporating Methods into Standards.....	449
4. More General Executive Branch Directives Aimed at Scientific Integrity and Transparency.....	451

* U.S. Environmental Protection Agency. The views in this article are held by the author and do not necessarily represent the views of the agency.

** Joe. A. Worsham Centennial Professor, University of Texas School of Law

*** Reader in Environmental Law, Corpus Christi College and Faculty of Law, University of Oxford.

II. WHY METHODS ARE INVISIBLE	452
A. Conventional Misunderstandings and Methodological Complacency Among Lawyers	452
B. Reinforcing Institutional Incentives	454
1. Science Advisory Committees	454
2. Judicial Review	456
3. The Larger Regulatory Process	458
III. HOW TO CONSIDER METHODS	459
A. Department of Interior's Use of Probabilistic Inference	460
B. EPA's Synthesis of Scientific Information on Air Pollutants	465
C. Encouraging Visible Methods by Adjusting Judicial Review	468
CONCLUSION	470

INTRODUCTION

In 2001, hours after the Ernsts dined at the Olive Garden restaurant where they first dated, Robert—exercise fanatic, triathlete, marathon runner—died of an apparent heart attack.¹ “It didn’t make sense to me that he could have died like that from a heart problem,” said his wife, Carol.² Scouring the Internet, she discovered that Vioxx, the painkilling drug her husband had started taking, might be linked to cardiac arrest.³ She filed suit against Merck & Co., makers of Vioxx. After the jury awarded Carol more than \$250 million in damages,⁴ Merck’s defense team insisted the company acted responsibly “from researching Vioxx prior to approval in clinical trials . . . to monitoring the medicine while it was on the market, to voluntarily withdrawing the medicine when it did.”⁵ Yet even with these precautions, a scientist at the Food and Drug Administration (FDA) estimated that Vioxx caused roughly 55,000 deaths *after* the FDA approved it for market.⁶

Unfortunately, the deaths from Vioxx are only one example of how drugs that have been approved by the FDA may nevertheless lead to unex-

1. Alex Berenson, *In First of Many Vioxx Cases, a Texas Widow Prepares to Take the Stand*, N.Y. TIMES, July 13, 2005, at C1.

2. Kevin McCoy, *Merck to Face First Vioxx Trial Before Texas Jury Next Month*, USA TODAY (June 29, 2005), http://usatoday30.usatoday.com/money/industries/health/drugs/2005-06-29-vioxx-cover-usat_x.htm.

3. *Id.*

4. The trial court’s judgment was reversed on appeal. *Merck v. Ernst*, 296 S.W.3d 81 (Tex. App. Houston 2009), *cert. denied*, 132 S. Ct. 1980 (2012).

5. Marc Kaufman, *Merck Found Liable in Vioxx Case*, WASH. POST, Aug. 20, 2005, at A01.

6. Gardiner Harris, *F.D.A. Failing in Drug Safety, Official Asserts*, N.Y. TIMES, Nov. 19, 2004, at A1 (based on Dr. David Graham’s testimony before the Senate Finance Committee, available at <http://www.finance.senate.gov/imo/media/doc/111804dgttest.pdf>).

pected deaths, sometimes at considerable levels. Indeed, researchers estimate that roughly one out of every five FDA-approved drugs are likely to be linked to serious adverse side effects,⁷ risks that were not detected during the \$1 billion-plus testing and oversight of each drug that is required for FDA approval.⁸ The conventional explanation offered for this alarming gap in public health protection is that the clinical trials cannot detect all of the potential adverse effects due to the small size of the tested populations.⁹ While this may be partially true, we maintain that the FDA's failure to flag the link between Vioxx and the risk of cardiac arrests stemmed from a fundamental misunderstanding of the methods of scientific inference, which might have been avoided if these methods had been more visible to the regulatory community—including agencies, regulated entities, non-governmental organizations who serve as watchdogs over regulations, and the courts.

Scientific inferences are theories about how the world works, which scientists formulate on the basis of their observations.¹⁰ Such inferences are inevitably replete with uncertainties. Alternative theories can explain the same set of observations. Moreover, observations can often be incomplete and can sometimes be imprecise or inaccurate. Scientists must therefore rely on various analytical methods to evaluate the concordance between observations and the competing theories used to explain them.¹¹ We argue that the FDA promulgated only one among these various methods as the evidentiary standard to evaluate drug risk. Once it did so, this method evolved into a regulatory process—unexamined, unquestioned, and for the most part, invisible to the regulatory community—regardless of whether

7. See, e.g., Karen E. Lasser et al., *Timing of New Black Box Warnings and Withdrawals for Prescription Medications*, 287 JAMA 2215, 2216 (2002) (estimating that 20 percent of all new drugs are likely to have serious or life-threatening adverse effects within the first twenty-five years of use that will either be unknown or undisclosed at the time of drug approval). It has also been estimated that 250,000 adverse drug experiences (ADEs), i.e., adverse events that occur when drugs are used in professional practice, are reported to the FDA each year. It has been further estimated that costs associated with ADE-related mortality and morbidity exceed \$75 billion annually and that ADEs are among the top ten causes of death. Syed Rizwanuddin Ahmad, *Adverse Drug Event Monitoring at the Food and Drug Administration*, 18 J. GEN. INTERNAL MED. 57, 57 (2003).

8. See, e.g., Steven M. Paul et al., *How to Improve R&D Productivity: The Pharmaceutical Industry's Grand Challenge*, 9 NATURE REVIEWS DRUG DISCOVERY 203, 205 (2010), available at <http://www.nature.com/nrd/journal/v9/n3/full/nrd3078.html> (discussing the high costs of drug testing).

9. See, e.g., Lasser et al., *supra* note 7, at 2218 (concluding that adverse effects appear for new drugs because “[p]remarketing drug trials are often underpowered to detect ADRs, and have limited follow-up”).

10. RAY HILBORN & MARC MANGEL, *THE ECOLOGICAL DETECTIVE: CONFRONTING MODELS WITH DATA* 12 (1997).

11. *Id.*

this particular method was suitable to the observations on hand.¹² Indeed, with the benefit of hindsight from Vioxx, Congress in 2007 passed legislation to amend the FDA's incomplete approach and to provide the FDA with the authority to establish the infrastructure needed to collect observations that lend themselves to a broader set of inferential methods.¹³

The problems with the FDA's drug approval program before the 2007 legislation, as well as other problems in areas of environmental and health regulation, arose from a very simple but unappreciated problem—the failure to make methods visible as part of the regulatory process. The visibility we prescribe amounts to little more than requiring agencies to explicate and justify the choice of one inferential model over another. Rarely do agencies compare, explain, or justify these inferential techniques when synthesizing the scientific evidence that serves as the basis for regulatory decisions. As the Vioxx case and many other unfortunate regulatory fallouts attest, however, the consequences of hidden inferential methods can be potentially significant. As long as the basis for science-intensive regulatory decisions is obscure, the potential for errors and incomplete analyses on fundamental decisions such as risk prevention remains unnecessarily high.

Visible methods open up an agency's decisionmaking to scrutiny, both internal and external, in a way that would not be possible otherwise. Once agency staff are expected to make their methods visible, such methods would not be made implicitly or by default, or without a careful consideration of alternatives. Perhaps more importantly, requiring methods to be visible would shift the courts' oversight role toward insisting on explanations rather than evaluating individual agency choices themselves, except with respect to their fit within the statutory frame. Finally, visible methods would refocus the full range of regulatory participants on the methodological choices available and the importance of selecting the best inferential methods for the task at hand. In short, making methods visible highlights the “means” of the decisionmaking or how inferences are derived from data, rather than on the outcomes or “ends” of regulatory deliberations.

While these methods of scientific inference may seem far outside the domain of most lawyers, we argue that the legal system in general and the courts in particular are partly to blame for the deficiencies in the agencies' lack of methodological transparency. The process of judicial review has not only been complicit in allowing the agencies' methods of inferring risks and synthesizing evidence to fall out of sight, but has actually aggravated the tendency by making methods largely irrelevant to the basic mechanisms

12. See *infra* Part I.

13. Food And Drug Administration Amendments Act of 2007, Pub. L. No. 110-85, 121 Stat. 823.

used to hold agencies accountable.¹⁴ What is worse, when decisionmakers do develop detailed, visible methods, the legal system provides little reward or encouragement for such efforts. If anything, contemporary statutory requirements and mechanisms for judicial review can actually serve to penalize this methodological candor through misguided second-guessing.¹⁵

Of even greater concern, there is virtually no debate within legal arenas about the nature and importance of methods of scientific inference. Few lawyers will have ever thought about the methods by which scientific findings are inferred from data.¹⁶ Lawyers have instead treated scientific analysis as a "truth machine"¹⁷ which produces answers but they have not felt the need to scrutinize how the machine *actually* works. Frameworks such as "risk assessment"¹⁸ might give the appearance of a method, but in fact are merely general decisionmaking frameworks derived from concerns over administrative legitimacy¹⁹ and imposed upon regulatory science as part of the administrative process. To use an analogy, it is as if the work of common law courts were simply understood in terms of the results of decided cases rather than the reasoning deployed by judges. To do this is to miss a significant aspect of the legal method; as lawyers, we not only focus on judicial method, but also see the visibility of such method as a virtue of judicial reasoning. A judgment that provides no reasoning or provides reasoning that is opaque is problematic, and there is a rich discourse about the nature and quality of legal methods.²⁰ All the same is true in relation to methods of inference in science.

Once the methods are made visible, scientists, policymakers, and interested participants will be forced to grapple with identifying the most robust

14. See *infra* Part II.

15. See *infra* Part II.

16. There are notable exceptions, however. See, e.g., Steve C. Gold, *A Fitting Vision of Science for the Courtroom*, 3 WAKE FOREST J.L. & POL'Y (forthcoming 2013) (Rutgers Sch. of Law-Newark, Paper No. 118, 2012) (arguing that in reviewing the admissibility of expert weight of the evidence testimony under *Daubert*, the First Circuit Court of Appeals in *Milward* rightly demanded rigorous explanations and methods to support the experts' inferences, rather than direct testing), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2101454.

17. Wendy Wagner, Elizabeth Fisher & Pasky Pascual, *Misunderstanding Models in Environmental and Health Regulation*, 18 N.Y.U. ENV. L.J. 293 *passim* (2010).

18. COMM. ON THE INST. MEANS FOR ASSESSMENT OF RISKS TO PUB. HEALTH, NAT'L RESEARCH COUNCIL, RISK ASSESSMENT IN THE FEDERAL GOVERNMENT: MANAGING THE PROCESS (1983), available at http://books.nap.edu/openbook.php?record_id=366&page=1.

19. ELIZABETH FISHER, RISK REGULATION AND ADMINISTRATIVE CONSTITUTIONALISM ch. 3 (2007).

20. E.g., RONALD DWORKIN, LAW'S EMPIRE (1991); MARTIN SHAPIRO, THE SUPREME COURT AND ADMINISTRATIVE AGENCIES (1968). Judicial reasoning has been a major scholarly focus in the area of risk regulation. See Richard B. Stewart, *The Role of Courts in Risk Management*, 16 ENVTL. L. REP. 10208 (1986); Patricia Wald, *Regulation at Risk: Are Courts Part of the Solution or Most of the Problem?*, 67 S. CAL. L. REV. 621, 636-39 (1994).

and comprehensive approaches to inference. While deconstruction and ends-oriented attacks against methods remain fair game, they become a much more difficult means of sabotage as long as critics are expected to point to uniformly superior approaches that could replace the methodological flaws they attack. By making methods visible, the regulatory apparatus also becomes more conditioned to finding ways to remain updated and to evolve with improved methods. The agencies' focus shifts towards developing the best methods, not simply developing scientific outputs that are presented strategically in ways intended to navigate around possible legal and political controversies and lines of attack.

We do not pretend that making methods visible is a magic wand solution to ossification,²¹ the manufacturing of scientific uncertainty,²² or analytical opportunism on the part of litigants, however.²³ By drawing the focus to how scientists draw inferences from the data, lawyers are focusing on the most fundamental element in the rigor of science-intensive administrative decisionmaking. While we could be accused of having little awareness of how science can be socially constructed and framed,²⁴ such a criticism neglects the crux of our argument; namely, that methods need to be made more visible in the regulatory process, not more authoritative.²⁵ As Bruno Latour has remarked, the key problem in contemporary science debates today is not so much that we have an "excessive confidence in ideological arguments posturing as matters of fact"—dismantling regulatory science seems to be routine—but that instead we have "an excessive distrust of good matters of fact disguised as bad ideological biases."²⁶ Making method visible is thus not about uncovering a 'truth machine' but about ensuring that a discourse takes place about what are the methods best suited for the particular factual questions that underpin a specific regulatory decision in an area in which there are limits to scientific knowledge.

Our overarching argument that risk regulation would be considerably improved—on all levels—by making methods visible is developed in three parts. Part I provides a very particularized discussion of methods, explain-

21. See Thomas O. McGarity, *Some Thoughts on "Deossifying" the Rulemaking Process*, 41 DUKE L.J. 1385, 1403 (1992).

22. See generally DAVID MICHAELS, DOUBT IS THEIR PRODUCT: HOW INDUSTRY'S ASSAULT ON SCIENCE THREATENS YOUR HEALTH (2008).

23. FISHER, *supra* note 19, at ch. 3.

24. The literature on this point is a rich one. See SHEILA JASANOFF, SCIENCE AT THE BAR: LAW, SCIENCE AND TECHNOLOGY IN AMERICA (1995); BRUNO LATOUR, SCIENCE IN ACTION (1987).

25. See FISHER, *supra* note 19; Pasky Pascual, *Wresting Environmental Decisions from an Uncertain World*, 35 ENVTL. L. REP. 10539 (2005); Wendy E. Wagner, *The Science Charade in Toxic Risk Regulation*, 95 COLUM. L. REV. 1613 (1995).

26. Bruno Latour, *Why Has Critique Run Out of Steam? From Matters of Fact to Matters of Concern*, 30 CRITICAL INQUIRY 225, 227 (2004).

ing what we mean by “methods”; of how scientists use a variety of methods to draw inferences from the available information; and of why scientific inferences can be justified only within the context of methods used to derive them. Many lawyers might be tempted to skip or skim this Part on the assumption that such scientific discussion is within neither their expertise nor domain.²⁷ We would strongly urge readers not to do this, however; this Part lays bare some of the gaps that can arise when the methods of analysis are invisible, as they were in relation to the FDA approval of Vioxx. Part I concludes by discussing why the invisibility of methods has led regulatory agencies down the path toward inconsistent analyses of pesticide risks and untenable decisions regarding a variety of standards, benchmarks, and regulatory pronouncements.

Part II considers why methods tend to be obscure in regulatory debates. As already noted above, we argue that currently very little method is actually made visible and that frameworks such as “risk assessment” and monikers such as “sound science” serve as effective black boxes that promote the invisibility of methods. This Part explores why such an inferior state of affairs is institutionally tolerated.

Part III considers how methods should be made visible and then provides several illustrations on how this might be done. We argue that such visibility can serve a range of ends, but the most significant is improving the quality of regulatory decisionmaking. Making methods visible is important because it is about making the reasoning process visible so it can be assessed. This is particularly so in administrative law because reasoning and its rigor has always been the subject’s historical focus. We conclude the third Part by suggesting how courts can enhance the visibility of methods by rewarding agencies for this work.

We should stress at the outset that this Article is part of a larger inquiry into the interface between science and law that we have been carrying out over the last few years.²⁸ That interface has largely been characterized as an uneasy and even unviable collaboration.²⁹ What we have been interested in doing is exploring how that relationship can be understood in a more constructive way. Making method visible is a necessary, albeit not the only step, needed to do this.

27. Elizabeth Fisher et al., *Understanding Environmental Models in Their Legal and Regulatory Context*, 22 J. ENVTL. L. 251 (2010) (critiquing that assumption).

28. See, e.g., Wagner et al., *supra* note 17.

29. See, e.g., Devra Lee Davis, *The Shotgun Wedding of Science and Law: Risk Assessment and Judicial Review*, 10 COLUM. J. ENVTL. L. 67, 68 (1985); Cary Coglianese & Gary E. Marchant, *Shifting Sands: The Limits of Science in Setting Risk Standards*, 152 U. PA. L. REV. 1255 (2004).

I. METHODS AND WHY THEY MATTER

Regulatory crises such as the Vioxx tragedies, along with other unexpected adverse events associated with approved drugs, result in part from a tacit selection of methods at a particular point in time that outlasts the appropriate use of those methods in consequent regulatory decisions. Before discussing how methods matter to regulatory outcomes, however, we first discuss what we mean by methods of inference. There are not only choices between methods of inference, but several layers at which those choices consistently arise. Each level of choice, moreover, can impact regulatory outcomes, potentially exponentially.

A. What Are the Methods of Scientific Inference?

On one level, the Vioxx crisis stemmed from questions over a scientific fact—was the drug factually safe or not? Such questions accompany virtually all major public health and safety regulations: Does chloroform in drinking water pose risks at any concentration above zero?³⁰ Does bisphenol A in baby bottles give rise to risks?³¹ Do greenhouse gases endanger public health and the environment?³² There has been much literature concerning how the question of safety is not just a factual question but also a value-based one.³³ That is true, but focusing on this ignores that underlying the factual aspects of these questions is a subtler, more fundamental scientific issue with significant, practical consequences; ultimately, disputes over facts have as much to do with disagreements over which methods establish the most believable and most scientifically trustworthy description of the data.³⁴

30. In *Chlorine Chemistry Council v. EPA*, 206 F.3d 1286 (D.C. Cir. 2000), the court remanded EPA's drinking water standard for chloroform of 0 parts per million because some evidence suggested there was a safe threshold for cancer risk.

31. The FDA ruled that baby bottles and children's drinking cups could no longer contain BPA, although it did so based on the industry's abandonment of BPA in the manufacture of those items. Indirect Food Additives: Polymers, 77 Fed. Reg. 41,899, 41,902 (July 17, 2012) (to be codified at 21 C.F.R. pt. 177). The FDA continues to study the safety risks of BPA. FOOD & DRUG ADMIN., BISPHENOL A (BPA): USE IN FOOD CONTACT APPLICATION, FDA.GOV, <http://www.fda.gov/NewsEvents/PublicHealthFocus/ucm064437.htm> (last updated Apr. 2, 2012).

32. See *Coalition for Responsible Regulation v. EPA*, 684 F.3d 102, 121 (D.C. Cir. 2012).

33. E.g., K.S. SHRADER-FRECHETTE, BURYING UNCERTAINTY: RISK AND THE CASE AGAINST THE GEOLOGICAL DISPOSAL OF NUCLEAR WASTE 27-102 (1993); Steven Rayner & Robin Cantor, *How Fair Is Safe Enough? The Cultural Approach to Societal Technology Choice*, 7 RISK ANALYSIS 3 (1987).

34. See M.P. Lynch, *Epistemic Circularity and Epistemic Incommensurability*, in SOCIAL EPISTEMOLOGY 262, 262-64 (A. Haddock et al. eds., 2010).

Every rigorous discipline has methods that practitioners use to evaluate the quality of work conducted within their domain. Law is no exception. In countries practicing civil law, judges reach their decisions by applying a code of legal principles to the litigation at hand. Judges operating within a system of common law analyze the case history of past judicial opinions to reason their way to decisions.³⁵ Without a jurisdiction's shared understanding of principles governing how to reach legal conclusions, there can be no agreement on whether judicial opinions should be trusted or believed.³⁶

While there has been a gradual acknowledgement within the law that scientific knowledge is distinguished by information that results from the deployment of a "scientific method,"³⁷ much less discussed are the methods themselves—the epistemic principles—through which scientists draw inferences from their data.³⁸ One example of such inferential methods is the "aspects," essentially qualitative causal assumptions, developed by Sir Austin Bradford Hill in 1965 to guide physicians in understanding the causal link between disease and environmental factors. Hill proposed several aspects for consideration: the strength or frequency of observed associations; consistency of association in varied circumstances; specificity of association; temporal relationship between disease and posited cause; the dose response curve between them; biological plausibility of the causal explanation; coherence of the explanation with aspects of the disease; experimental data; and existence of analogous causal relationships.³⁹ Hill's aspects have stood the test of time. They have been widely accepted by epidemiologists and toxicologists.⁴⁰ They have also been used in numerous judicial opinions⁴¹ and agency regulations,⁴² despite the fact that no

35. BARRON'S LAW DICTIONARY 96 (6th ed. 2010).

36. This of course raises interesting and important questions about how foreign case law is used by courts. See Stephen Yeazell, *When and How U.S. Courts Should Cite Foreign Law*, 26 CONST. COMMENT. 59 (2009); Christopher McCrudden, *Common Law of Human Rights?: Transnational Judicial Conversations on Constitutional Rights*, 20 OXFORD J. LEGAL STUD. 499 (2000).

37. See, e.g., *Daubert v. Merrell Dow Pharm.*, 509 U.S. 579 (1993); *Kitzmiller v. Dover Area Sch. Dist.*, 400 F. Supp. 2d 707, 736 (M.D. Pa. 2005) (referring broadly to a method in which explanations of natural phenomena are inferred from "confirmable data—the results obtained through observations and experiments that can be substantiated by other scientists").

38. By "inference," we mean conclusions about what the data are saying. All inferential methods are based on "first principles" which cannot be justified by *a priori* reasons. The global community of scientists has gravitated towards these principles because they have proven to be useful, to be relatively objective, to be rational, and they make reasoning visible.

39. A.B. Hill, *The Environment and Disease: Association or Causation?*, 58 PROC. ROYAL SOC'Y MED. 295, 295–99 (1965).

40. Gerard Swaen & Ludovic van Amelsvoort, *A Weight of Evidence Approach to Causal Inference*, 62 J. CLINICAL EPIDEMIOLOGY 270, 270 (2009).

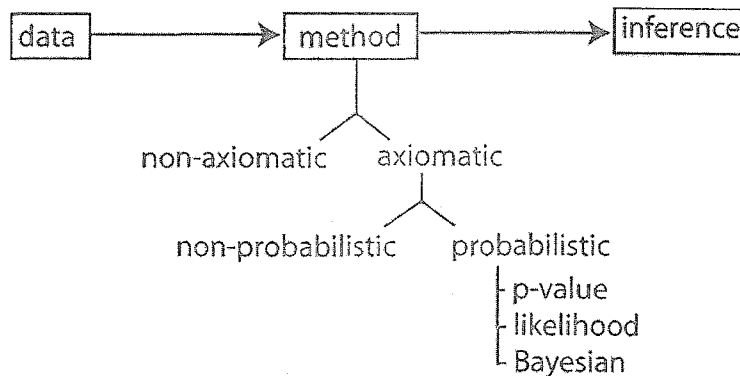
41. E.g., *Milward v. Acuity Specialty Prods. Grp., Inc.*, 639 F.3d 11, 17 (1st Cir. 2011).

algorithm exists to apply these factors, which Hill conceded were non-exhaustive, unranked, and defeasible.

To formulate inferences based on quantitative methods, the scientific community has developed ways to build on qualitative principles by using computations or “algorithms”—logical, sequential systems of computations—that are grounded in axioms. Axioms are epistemic first principles; they can only be assumed to be true. By using these axioms, scientists have developed computational approaches to extract inferences about data. The appeal of these inferential methods lies in their universality, and replicability and the way in which they make reasoning visible. Any scientist anywhere in the world who applies the same method to the same data should generate the same set of inferences.

Foremost among these methods are those based on the axioms of probability. Probabilistic inference is a discipline unto itself and we can make no pretense here of providing a comprehensive overview of it. However, it is a predominant approach for scientific inference. Appreciating its incomplete features and even its misuse is key to understanding not only how the FDA misapprehended the risks from Vioxx, but also how courts misconstrue the weight of scientific evidence. For this reason, in the next few sections, we will provide intuitive explanations of the logic of probabilistic inference.

FIGURE 1. VARIETY OF METHODS FOR SCIENTIFIC INFERENCE



Data collected may not lend themselves to quantitative analysis. In such cases, scientists reach inferences based on non-axiomatic principles, such as the factors identified by Hill. For the most part, when the data justify their use, scientists use inferential methods based on the axioms of probability.

Figure 1 summarizes the variety of inferential methods used by the scientific community. We broadly categorize these methods into those based on axiomatic systems and those that are not. As an example of the latter,

42. *E.g.*, National Ambient Air Quality Standards for Ozone, 73 Fed. Reg. 16,436, 16,456 (March 27, 2008) (to be codified at 40 C.F.R. pts. 50, 58).

scientists use qualitative factors such as those used by Hill to infer relationships between environmental and health harms and their causes. As to the former, scientists assume axioms so that they may formally apply computational approaches and quantitatively describe these relationships. Foremost among these axiom-based methods is probabilistic inference.⁴³ We postpone our discussion of probabilistic inference until Part II. For now, we only note the existence of other axiomatic systems that are not probabilistic. For example, some commentators argue that long-term, complex systems, such as climate change and radioactive waste disposal, are so uncertain that their risks can only be estimated based on opinions elicited from experts in specific fields. These opinions are translated into numbers that can be synthesized and computed based on axioms other than those governing probability.⁴⁴

Three insights follow from Figure 1. First and most obvious, the methods of scientific inference are varied. Yet, despite this variety, only *one* type of method seems to garner the disproportionate attention of the regulatory process. This is the method of *p*-values, a type of probabilistic inference. In the words of one commentator, it is “the traditional measure of evidence, which I think is baked into every brick of regulatory buildings.”⁴⁵ It is the method alluded to when there is an insistence upon “statistically significant” results before giving credence to scientific evidence.⁴⁶

Second, the methods outlined in Figure 1 are not mutually exclusive. Indeed, when scientists rely on multiple studies, conducted with multiple approaches, and the inferences derived from these studies converge around a coherent explanation of the relationship between harm and its suspected cause, the convergence is an indication of the strength of the evidence.

Third, to draw reliable inferences, application of a method's axioms should be appropriate to the data. As we discuss next, ignoring this truism contributed to the FDA's undervaluation of the risks from Vioxx.

43. For technical discussions of the variety of axiom-based, inferential methods, see George J. Klir & Richard M. Smith, *On Measuring Uncertainty and Uncertainty-based Information: Recent Developments*, 32 ANNALS OF MATHEMATICS AND ARTIFICIAL INTELLIGENCE 5 (2001) and Lofti A. Zadeh, *Toward a Generalized Theory of Uncertainty (GTU)—an Outline*, 172 INFO. SCI. 1 (2005).

44. Scott Ferson et al., *Summary from the Epistemic Uncertainty Workshop: Consensus amid Diversity*, 85 RELIABILITY ENGINEERING & SYS. SAFETY 355 (2004); Jon C. Helton et al., *Representation of Analysis Results Involving Aleatory and Epistemic Uncertainty*, 39 INT'L J. GEN. SYS. 605 (2010).

45. Steven N. Goodman, *Introduction to Bayesian Methods I: Measuring the Strength of Evidence*, 2 CLINICAL TRIALS 282, 284 (2005).

46. See, e.g., *Milward v. Acuity Specialty Prods. Grp., Inc.*, 664 F. Supp. 2d 137, 148–49 (D. Mass 2009). This decision was overturned on appeal, when the higher court acknowledged a broader set of inferential methods should be acceptable in court. *Milward*, 639 F.3d at 23–25.

B. How Was FDA Constrained to Rely on the Method of P-values as Its Sole Method of Making Inferences?

We return to the fate of Robert Ernst and the larger controversy surrounding the FDA's approval of Vioxx. To the extent that weaknesses arose in the FDA's program, they are suggestive of broader maladies across agencies. Indeed, in the case of the FDA and possibly other agencies as well, the narrow inferential frame was primarily if not exclusively the result of legislative design rather than agency choice.

After considerable trial and error in drug regulation,⁴⁷ Congress passed the 1962 amendments to the legislation governing drug regulation, which mandated that the FDA assess whether a drug was effective for its intended use based on "substantial evidence" from "adequate and well-controlled investigations."⁴⁸ This was the legislation in force during the FDA's approval of Vioxx, and the FDA interpreted this statute to mean that a regulatory decision on drug effectiveness must be based on *randomized, replicated, controlled, clinical trials (RCTs)*.⁴⁹ In essence, RCTs serve as the idealized, traditional experiment in which experimental conditions are held constant and homogeneous, except for the causal factors under investigation (in this case, different levels of drug dose). It is this experimental design that justifies using the method of *p*-values.

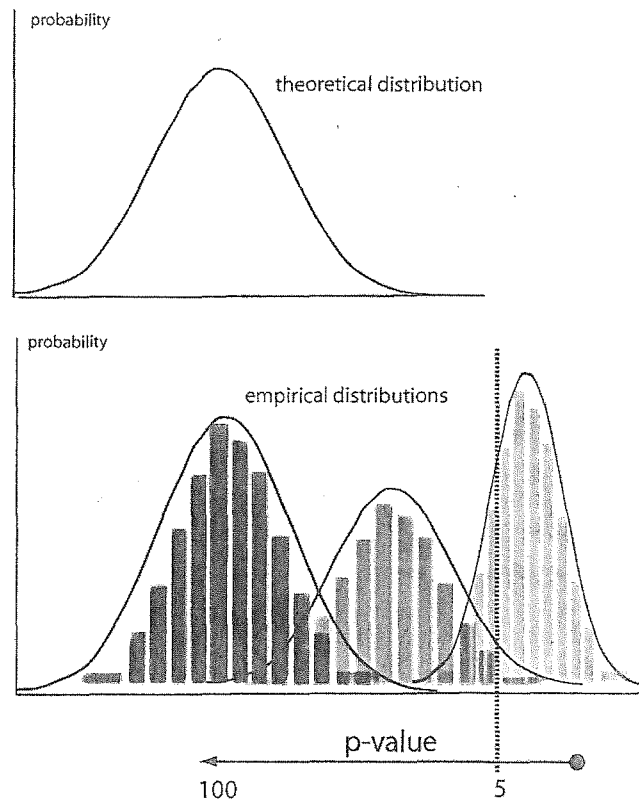
The axioms of probability imply that (1) patterns in data about the natural world can be approximated by mathematical forms—probability distributions, of which the so-called "bell curve" is archetypical; and (2) these forms can be manipulated computationally in order to evaluate hypotheses about how the world operates.⁵⁰ The method of *p*-values builds on these axioms, as we explain in Figure 2.

47. For some of this history see CHARLES O. JACKSON, *FOOD AND DRUG LEGISLATION IN THE NEW DEAL* (1970); GUSTAVUS A. WEBER, *THE FOOD, DRUG AND INSECTICIDE ADMINISTRATION: ITS HISTORY, ACTIVITIES, AND ORGANIZATION* (1928); David F. Cavers, *The Food, Drug and Cosmetic Act of 1938: Its Legislative History and Its Substantive Contents*, 6 L. & CONTEMP. PROBS. 2 (1939).

48. Drug Amendments of 1962, Pub. L. No. 87-781, § 102(d), 76 Stat. 780, 781 (codified at 21 U.S.C. § 355(d) (2006)).

49. 21 C.F.R. § 314.126 (2012). The initial rules adopted by FDA have evolved over time, as described in Russell Katz, *FDA: Evidentiary Standards for Drug Development and Approval*, 1 NEURORX 307, 307-09 (2004).

50. See SCOTT M. LYNCH, *INTRODUCTION TO APPLIED BAYESIAN STATISTICS AND ESTIMATION FOR SOCIAL SCIENTISTS* (2007) (discussing probability axioms, distributions, and their applications).

FIGURE 2. THE METHOD OF P -VALUES

Inferential methods based on p -values follow the logic of proof by contradiction.

- We first set up experimental groups. These are completely homogeneous, except for varying levels of the causal factor being investigated. We initially assume there is no causal relationship. Therefore, we expect data from the experimental groups to conform to a theoretical distribution showing only some random pattern of variability (e.g., top graph).
- After running the experiment, we estimate the distribution of the observed data (e.g., bottom graph shows three hypothetical empirical distributions).
- We expect some differences between the theoretical and empirical distributions because of happenstance. However, the difference might be so extreme that we cannot attribute to chance alone the difference between the theoretical and empirical distributions.
- The p -value measures this difference. It is the probability of observing the empirical distribution, assuming the theoretical, random distribution is true. If this probability is very low (by convention, less than 5%), then we infer that causal factors generated the difference. In the bottom graph, it is improbable that chance alone generated the rightmost distribution; causal factors probably influenced that distribution.

When scientists speak of “statistically significant” results, they typically mean that the experimental results generated a p -value less than 5%. It is important for lawyers and judges to understand that this notion of statistical significance is meaningful *only within the context of the ideal experimental paradigm*, in which comparisons are made among groups that are homoge-

neous but for the causal factor being investigated.⁵¹ As explained in Figure 2, it is this experimental design that justifies the initial assumption of a random, theoretical distribution, which serves as the basis for the p -value.

But the very steps taken to preserve the experimental conditions demanded by the method of p -values ultimately led to the unregulated risks from Vioxx. To preserve homogeneity in the RCTs for Vioxx, Merck excluded older patients with previous cardiovascular disease.⁵² Yet, it was precisely this cohort that was later shown to be at greatest risk.⁵³ Moreover, some of the variability causing differences in the response to Vioxx was observable only at the gene level.⁵⁴ This complicated the extrapolation of RCT results to the general population. Finally, the incidence of heart disease from Vioxx in the greater population may have been so rare as not to have generated the low p -values needed for statistical significance. But the rarity of the event did not minimize the consequences of risk, as Robert Ernst's tragedy demonstrated.

Prior to the Vioxx incident, the fact that the FDA's methods for drug approval were based on a limited inferential method was wholly invisible to most policymakers and many interested parties—including drug manufacturers—that, if asked, would likely have opted for a more comprehensive or a different form of assessment. Particularly since the FDA did not view itself as having much legislative choice, the methods it followed had well-known blind spots and related limitations that seemed to be wholly ignored. Given the general invisibility of methods for inference across all agencies, this is not surprising. Indeed, having a debate about the strengths and weaknesses of inferential method, when such method is not visible and few regulatory actors recognize its importance is nigh on impossible.

In hindsight, it is clear that the FDA need not have used an exclusive inferential method to evaluate drug risk. As shown in Figure 1, scientific inference typically proceeds down multiple paths. It was only after considerable oversight hearings, scientific review, and agency self-reflection that a more comprehensive approach to scientific inference came to light for the FDA. Shortly after the Vioxx recall, the National Academy of Sciences' (NAS) Institute of Medicine (IOM) issued a report clearly stating what others had been saying for some time: the FDA's pre- and post-approval

51. See Steven N. Goodman, *Toward Evidence-Based Medical Statistics. 1: The P Value Fallacy*, 130 ANNALS INTERNAL MED. 995 (1999).

52. James M. Ritter et al., *Precaution, Cyclooxygenase Inhibition, and Cardiovascular Risk*, 30 TRENDS PHARMACOLOGICAL SCI. 503, 504 (2009).

53. *Id.*

54. Christine G. St Germaine et al., *Genetic Polymorphisms and the Cardiovascular Risk of Non-Steroidal Anti-Inflammatory Drugs*, 105 AM. J. CARDIOLOGY 1740, 1740, 1743–44 (2010).

practices were unlikely to detect rare but serious drug risks.⁵⁵ Pre-approval, RCTs simply did not generate all the information needed to assess risks that arise when the general population is exposed to a drug.⁵⁶ Post-approval, the FDA did not possess the statutory authorities needed to implement a nationwide system to gather this information.⁵⁷ Indeed, the terms pre- and post-approval (or their synonyms, pre- and post-market) were not a useful construct to understand drug risks. In its stead, the IOM report advocated assessing safety over a drug's life cycle, in which data were to be continuously gathered from multiple sources for ongoing analyses.⁵⁸

In response to recommendations such as these, in 2007 Congress passed a statute that attempted to expand the FDA's evidentiary base and to encourage the FDA to adopt additional methods of inference to help identify risks that might go undetected by traditional drug testing. First, Congress directed the FDA to establish a network of data systems to integrate any and all information to evaluate drug risks.⁵⁹ Second, Congress provided the FDA with new, extensive authorities to require post-market submission of this information.⁶⁰ In interpreting these authorities, the FDA has stated that it can request this information if it will help assess serious drug risks, if it is unavailable in the FDA's network of data systems, and if the request is based on scientific data the agency deems appropriate.⁶¹ The types of information that the FDA can request from drug manufacturers are far-ranging: observational epidemiologic studies; electronic medical records and administrative health care claims; meta-analyses based on clinical trials or observational studies; and *in vitro* and *in vivo* laboratory studies involving animals.⁶² In brief, the evidentiary paradigm for the FDA's drug evaluation now goes beyond any single study or RCT, however well designed, and integrates evidence generated across multiple investigations.

The FDA's 2007 legislation was a necessary, but far from sufficient, step toward coherent, science-based decisions. It prompted the FDA to

55. See COMM. ON THE ASSESSMENT OF THE U.S. DRUG SAFETY SYSTEM, INST. OF MED. OF THE NAT'L ACADS., *THE FUTURE OF DRUG SAFETY: PROMOTING AND PROTECTING THE HEALTH OF THE PUBLIC* 37-38, 106-09 (A. Baciú et al., eds., 2007).

56. *Id.* at 37.

57. *Id.* at 153.

58. See *id.* at 169.

59. Food and Drug Administration Amendments Act of 2007, Pub. L. No. 110-85, § 905, 121 Stat. 823, 944 (codified at 21 U.S.C. § 355(k)(3)-(4) (2006)).

60. *Id.* § 901(a), 121 Stat. at 922-26 (codified at 21 U.S.C. § 355(o)-(p)).

61. FOOD & DRUG ADMIN., U.S. DEP'T OF HEALTH & HUMAN SERVS., GUIDANCE FOR INDUSTRY: POSTMARKETING STUDIES AND CLINICAL TRIALS—IMPLEMENTATION OF SECTION 505(O)(3) OF THE FEDERAL FOOD, DRUG, AND COSMETIC ACT 6 (2011), available at <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM172001.pdf>.

62. *Id.* at 7-10.

assess drug risk using inferential methods beyond RCTs and *p*-values. However, as we stated earlier, scientific inferences are unavoidably uncertain because observations are imperfect and because alternative theories can be used to explain identical sets of observations.⁶³ The most challenging aspect of scientific inference—the challenge that lies at the intersection of law and science—is to determine which combination of data and methods best contributes to the weight of evidence supporting one inference versus other competing inferences.⁶⁴

As we discussed earlier, the scientific community relies on multiple inferential methods, with each method resting on its own particular set of epistemic foundations.⁶⁵ Congress has neither the omniscience nor the competence to prescribe the appropriate method for every possible scenario requiring an evaluation of public health or environmental risks.

What we suggest therefore is that, when agencies issue regulatory decisions based on science, they must use inferential methods that are both pluralistic and visible. By “pluralistic,” we mean that agencies should openly recognize that their methods will vary, depending on the type of observations they are using to draw their inferences. By “visible,” we argue that agencies should explain why they have chosen particular methods and why these methods lead to the inferences being proposed by the agencies.⁶⁶ When agencies use pluralistic and visible methods, the regulatory community is put into a better position to safeguard the rationality and thus the legality of regulatory decisions.

Embracing a broader understanding of scientific inference⁶⁷ need not lead us to the shoals of postmodern, or even post-normal,⁶⁸ scientific relativism. The various methods summarized in Figure 1 must still be bound by the admonition raised by Latour that while historically we have had “excessive confidence in ideological arguments posturing as matters of fact,” scholars and regulatory actors now tend to have “an excessive distrust of good matters of fact disguised as bad ideological biases.”⁶⁹ As that is the

63. See *supra* text accompanying note 11.

64. See Mark L. Taper & Subhash R. Lele, *The Nature of Scientific Evidence: A Forward Looking Synthesis*, in *THE NATURE OF SCIENTIFIC EVIDENCE: STATISTICAL, PHILOSOPHICAL AND EMPIRICAL CONSIDERATIONS* 527, 528–29 (Mark L. Taper & Subhash R. Lele eds., 2004).

65. See *supra* text accompanying note 45.

66. See *infra* text accompanying note 192.

67. Justice Sotomayor, writing for a unanimous Supreme Court, evinced such an understanding, saying scientists do “not limit the data they consider to the results of randomized clinical trials or to statistically significant evidence.” See *Matrixx Initiatives Inc. v. Siracusano*, 131 S.Ct. 1309, 1319–20 (2011).

68. See Silvio O. Funtowicz & Jerome R. Ravetz, *Science for the Post-Normal Age*, 25 *FUTURES* 739, 739–40 (1993).

69. See *supra* note 26 and accompanying text.

case, we need to focus not on whether something is true, but rather whether it is well or badly constructed.⁷⁰ Are the inferences drawn by any particular inferential method, when set against epistemic first principles, *well* or *badly* constructed? These questions can only be answered if the methods are visible to the regulatory community.

C. Invisible Methods in Practice

The consequence of invisible methods was manifested in the Vioxx tragedies, but the problem and the resulting confusion and conflict run throughout a number of regulatory processes. This Section explores how the agencies' failure to explicate methods of inference for their regulatory decisions can lead to confusion and controversy.

1. Incorporating Methods into Pesticide Registration and Species Protection

An ongoing conflict arising at the intersection of pesticide registrations and endangered species protection offers a particularly salient example of the regulatory conflict that can result from the agencies' failure to make their methods visible.⁷¹ Currently the U.S. Fish and Wildlife Service (FWS) and the Environmental Protection Agency (EPA) reach very different conclusions from the data about the potential adverse impacts of pesticides, as illustrated in the text box below. Although their methods are invisible, lurking behind the agencies' divergent choices of assumptions and models are very different statutory instructions for assessing risks. The FWS is tasked with preventing the extinction of endangered species, and when a species may be adversely affected by a federal activity, the Endangered Species Act (ESA) requires the FWS to use the "best available" evidence in a way that gives the endangered species the benefit of the doubt.⁷² By contrast, in its regulatory assessment of a pesticide registration, the EPA is required to balance the benefits of a pesticide against its costs to human health and environment.⁷³ This net balancing produces a much more open-ended framework that does not afford species the benefit of the doubt. Instead, the species' risks are compared against the benefits of the pesticide.

70. Bruno Latour, *An Attempt at a "Compositionist Manifesto"*, 41 *NEW LITERARY HIST.* 471, 474 (2010).

71. The nature of this controversy is summarized in *BD. ON ENVTL. STUD. & TOXICOLOGY, NAT'L ACAD. OF SCI., STATEMENT OF TASK: ECOLOGICAL RISK ASSESSMENT UNDER FIFRA AND ESA*, <http://dels.nas.edu/Study-In-Progress/Ecological-Risk-Assessment-Under-FIFRA/DELS-BEST-11-01> (last visited Jan. 10, 2013).

72. 16 U.S.C. § 1533(b) (2006).

73. EPA must ensure that the pesticide does not present "unreasonable adverse effects on the environment." 7 U.S.C. § 136a(a) (2006).

FIGURE 3. COMPARING FWS VS. EPA JUDGMENTS ON PESTICIDE RISKS TO ENDANGERED SPECIES⁷⁴

<i>Questions Arising in the Scientific Analysis</i>	<i>FWS's Answers</i>	<i>EPA's Answers</i>
Should a study with methodological problems be excluded from the analysis? (e.g., what is the definition of "best available science"?)	Not if part of the study does not suffer from the methodological problems and the findings of that part of the study suggest risks to endangered species.	Yes. Standard exclusion criteria exclude studies that have methodological flaws that cause the studies to be unreliable.
What types of endpoints in the studies ⁷⁵ should be considered?	Sub-lethal, indirect and cumulative effects on species must be considered.	Only endpoints that can be measured with some precision can be included in the analysis.
How should the limited research on chemical mixtures be integrated into the analysis?	The effects of chemical mixtures, as well as inactive ingredients, are critical to an assessment of risks to a species.	There is so much variation in mixtures that they cannot be included in a reliable model.
What types of assumptions should be included in the models when there is uncertainty about real world conditions?	Liberal spray drift ⁷⁶ assumptions must be factored into an exposure model.	Reasonable spray drift assumptions should be factored into an exposure model.
How should the available on species' range be integrated into the analysis?	The species' range should be measured by assuming the most expansive range.	Population models need to adopt reasonable assumptions and require documentation for all assumptions.

As Figure 3 reveals, there are important judgments at each point in the agencies' synthesis of the literature on pesticide risks to endangered species. At the first step, the agency must determine which of the existing studies inform the regulatory project and which do not. While one might imagine

74. These differences are drawn largely from *BD. ON ENVTL. STUD. & TOXICOLOGY*, *supra* note 71; from letters from EPA to NMFS regarding draft biological opinions on various pesticide decisions, see Letter from Debra Edward, Dir., Office of Pesticide Programs, to James H. Lecky, Dir., Office of Protected Res. 3-4 (Sept. 15, 2008), *available at* <http://www.epa.gov/espp/litstatus/effects/epa-to-nmfs.pdf>; Letter from Richard P. Keigwin, Jr., Dir., Special Review and Reregistration Div., to James H. Lecky, Dir., Office of Protected Res. 2 (Sept. 10, 2009), *available at* <http://www.epa.gov/espp/litstatus/11-18-08-nmfs-biop.pdf>; and from Telephone Interview conducted by Wendy Wagner with anonymous FWS Staff, Endangered Species Program (Jan. 26, 2012).

75. An endpoint is the adverse effect that a researcher measures in a toxicity study. Mortality is one of the most straightforward endpoints. Other endpoints include various measures of neurological effects (e.g., spontaneous locomotion of a mouse in an open field), tumors (e.g., benign and malignant), reproductive and development effects (e.g., brain weights of offspring at birth), etc. The challenge in toxicology is identifying one or more endpoints for a study that can be measured reliably. Behavioral change in animals, for example, is a much more difficult endpoint to measure as compared with mortality.

76. Spray drift refers to how far the pesticide sprays into the environment (and beyond the target) when it is applied. Spray drift is affected by a number of factors, including the contents of the pesticide product, its method of application, and wind speed.

that generic "exclusion/inclusion" criteria could be designed to sort out the available research, even decisions about how to use the literature depend on whether the agency seeks to afford every benefit of the doubt to the species or instead simply to produce a replicable, "mean" answer to a question. Choices also arise in identifying the parameters that will be used in a model. For example, what effects should be considered in predicting adverse impacts (e.g., sub-lethal effects or easily measured mortality) and what pesticides should be included (e.g., the entire chemical mix or one pesticide at a time)? Choices arise again in determining how to account for various scenarios, such as assumptions regarding spray drift, species' range, and even the misuse of pesticides during application. All of these decisions are informed by scientific and technical judgments about plausible options, yet none is resolved by them. While Figure 3 extracts only a handful of these choices, in regulatory assessments there are often dozens. According to one classic NAS report, often as many as fifty significant choices can punctuate any given effort to characterize the risks of a product.⁷⁷

Although they are invisible, the agencies' choice of methods profoundly affects their ultimate decisions. The EPA's approach to synthesizing the literature often allows pesticide products to remain on the market with only limited restrictions; the FWS's synthesis of the same scientific research, by contrast, leads to the opposite outcome.⁷⁸ The resulting battles between the two agencies have sparked protracted and extensive litigation by interest

77. COMM. ON THE INSTITUTIONAL MEANS FOR ASSESSMENT OF RISKS TO PUB. HEALTH, *supra* note 18, at 29–33.

78. A review of the correspondence to date between FWS and EPA on individual pesticide decisions reveals credible and difficult technical disagreements over critical judgments that should be used in making these assessments, with resulting implications for the ultimate pesticide licensing decisions. For example, FWS questions the viability of models used by EPA to estimate risks to endangered species or to calculate their ranges, *see* Letter from Marjorie Nelson, Chief, Branch of Consultation and HCPs, FWS, to Arthur-Jean B. Williams, Assoc. Dir., Envtl. Fate and Effects Div., Office of Pesticide Programs, EPA (Feb. 11, 2008), *available at* <http://www.epa.gov/oppfead1/endanger/litstatus/effects/atrazine/2008/fws-nonconcur.pdf>, and EPA questions the basis for various conclusions in the Biological Opinion about adverse effects resulting from the use of pesticide products. *See, e.g.*, Letter from Stephen Bradbury, Dir., Office of Pesticide Programs, EPA, to James H. Lecky, Dir., Office of Protected Res., NMFS, (Aug. 19, 2010), *available at* <http://www.epa.gov/oppfead1/endanger/litstatus/effects/final-biop-ltr.pdf>; Letter from Stephen Bradbury, Dir., Office of Pesticide Programs, EPA, to James H. Lecky, Dir., Office of Protected Res., NMFS (June 14, 2011), *available at* <http://www.epa.gov/oppfead1/endanger/litstatus/nmfs-draft-4-1comment.pdf>. Some progress has been made in bridging these different risk assessment approaches, however. In at least one set of pesticide reviews, EPA adjusted its registration requirements in accord with these agencies' biological opinions. *See* Letter from Richard Keigwin, Special Review and Reregistration Div., EPA, to James H. Lecky, Dir., Office of Protected Res., NMFS (Sept. 10, 2009), *available at* <http://www.epa.gov/oppfead1/endanger/litstatus/11-18-08-nmfs-biop.pdf>.

groups.⁷⁹ Ultimately, the two agencies commissioned the NAS to serve as referee on the best way to synthesize the evidence at the intersection of pesticide approvals and endangered species protections.⁸⁰ One of the first tasks of the NAS will be to draw out the various methodological steps of the analysis and then suggest how the agencies should best approach them, an exercise that ideally will produce a coherent and visible approach to merging the mandates.

The material above demonstrates that methods of inference and how they relate to the statutory mandate are the source of the conflict, not differences in the underlying evidence or even the basic agency scientific assessments of that evidence. Specifically, because the methods that the agencies used to reach their judgments were obscured, the clashes appeared both more inconsistent and irreconcilable than they actually were. It is only when the agencies unpack their inference methods and other assumptions that the agencies' analysis becomes accessible and can be compared across different programs that necessarily, by statute, deploy very different weighting factors for public policy. Many lawyers and policymakers, who have largely understood the scientific aspects of decisionmaking as a black box generating answers for the regulatory process, do not appreciate that fact, however.⁸¹ That gap in understanding only serves to further heighten the conflict and leave the core problem unresolved.

2. Incorporating Methods into How Categories of Regulated Industries Are Defined

Another, quite different example of the invisibility of methods in regulatory decisionmaking is the approach that agencies, primarily the EPA and OSHA, use to set technology-based standards for pollution control in air and water.⁸² The EPA in particular is instructed to identify a category of affected industries and to locate the best available pollution controls from within that set.⁸³

79. See, e.g., PESTICIDES: ENDANGERED SPECIES PROTECTION PROGRAM, EPA, <http://www.epa.gov/espp/litstatus/eslitig.htm> (last updated May 9, 2012) (describing the ongoing litigation).

80. See the agencies' charge to the NAS Committee examining ecological risk assessment under FIFRA and ESA, BD. ON ENVTL. STUD. & TOXICOLOGY, NAT'L ACAD. OF SCI., *supra* note 71.

81. A longer analysis of this problem can be found in Wagner et al., *supra* note 17.

82. See generally Wendy E. Wagner, *The Triumph of Technology-Based Standards*, 2000 U. ILL. L. REV. 83 (2000).

83. E.g., 42 U.S.C. § 7412(d)(3) (2006) (requiring that emissions from existing plants meet at least "the average emission limitation achieved by the best performing 12 percent of the existing sources").

Some industry groups discovered that one way to coax the EPA to lower the standards is to subdivide the industry into smaller and smaller categories. Steel manufacturers writ large might be required to install equipment to reduce pollution equivalent to that used by the top 5% of the performers in that sector; but if the EPA re-categorizes steel manufacturers into seven different subcategories, most of the top 5% performers in the subcategories may be much less successful in reducing pollution and the resulting pollution control standards can be significantly compromised in some and perhaps many of these subcategories.⁸⁴

This critical methodological step—identifying how the agency defined the larger set of affected industries—is rarely articulated or justified by the agency, however.⁸⁵ As a result, the standards are largely immune to scrutiny.

3. Incorporating Methods into Standards

Other methodological choices used by agencies have similarly remained obscure in ways that undermine the accountability of the regulatory programs. One example is the EPA's routine use of average adult susceptibilities to individual toxins to estimate mean effects of a pollutant on human health.⁸⁶ Such a methodological placeholder ignores synergistic effects and hot spots; sub-populations of extra-sensitive persons; and downplays the well-known added susceptibility of the children and elderly.⁸⁷ While gradually this

84. See, e.g., National Emission Standards for Hazardous Air Pollutants from the Portland Cement Manufacturing Industry, 74 Fed. Reg. 21,136, 21,140, 21,144, 21,148 (proposed May 6, 2009) (to be codified at 40 C.F.R. pts. 60, 63) (discussing the permissibility and advantages of subcategorizing industries to provide for higher emission standards for some groups of industry and avoid shutdowns that might otherwise result from a single emission standard, and citing Judge Williams's concurrence as endorsement of this approach).

85. One example is the deeply buried discussion of the authority to subcategorize industries to set the standards and the economic advantages to this technique in a court case where this practice was challenged. *Sierra Club v. EPA*, 479 F.3d 875, 885 (D.C. Cir. 2007) (“[Although] authority to generate subcategories is obviously not unqualified . . . one legitimate basis for creating additional subcategories must be the interest in keeping the relation between ‘achieved’ and ‘achievable’ in accord with common sense and the reasonable meaning of the statute.”) (Williams, J., concurring).

86. See, e.g., COMM. ON RISK ASSESSMENT OF HAZARDOUS AIR POLLUTANTS, NAT'L RESEARCH COUNCIL, SCIENCE AND JUDGMENT IN RISK ASSESSMENT ch. 11 (1994).

87. For an illustration in a research study for the importance of these variables, see Philip J. Landrigan et al., *Pesticides and Inner-City Children: Exposures, Risks, and Prevention*, 107 ENVTL. HEALTH PERSP. SUPPLEMENT 431 (1999) (spotlighting through a research study the unaccounted for synergistic effects, high levels of exposure that are unexpected, and increased susceptibilities of children, all in a single modest study measuring pesticide exposures to children in the inner-city).

methodological step is being examined in more detail,⁸⁸ retrospective adjustments may be complicated as a result of the invisible methods.

In a recent review of the EPA's formaldehyde risk assessment, the EPA was also taken to task for obscuring its major methodological assumptions in synthesizing the evidence and reaching key conclusions. Specifically, the panel observed that it was difficult to understand the EPA's assumptions and analysis on a number of points.⁸⁹ Indeed, the panel observed, these

[p]roblems with clarity and transparency of the methods appear to be a repeating theme over the years, even though the documents appear to have grown considerably in length. In the roughly 1,000-page draft reviewed by the present committee, little beyond a brief introductory chapter could be found on the methods for conducting the assessment.⁹⁰

The recommendations of the panel, consistent with our argument, urged the EPA to articulate its methods more completely and accessibly.⁹¹ For example, the panel recommended that the EPA should describe

more fully the methods of the assessment, including a description of search strategies used to identify studies with the exclusion and inclusion criteria clearly articulated and a better description of the outcomes of the searches (a model for displaying the results of literature searches is provided later in this chapter) and clear descriptions of the weight of evidence approaches used for the various noncancer outcomes.⁹²

Even monetizing the impacts of pollutants and other stresses on public health and the environment under Executive Order No. 12,866 suffers from methodological black boxes.⁹³ A focus on numbers and ultimate bright-line determinations of economic impact, without attention to developing an explicit discourse about rigorous methods for how these estimates can be developed, has led to analyses that appear more geared toward insulating the agency from litigation than advancing an understanding of the costs and

88. For EPA's efforts to develop supplemental guidelines for estimating cancer susceptibilities early in life, see EPA, PUB. NO. 630/R-03/003F, SUPPLEMENTAL GUIDANCE FOR ASSESSING SUSCEPTIBILITY FROM EARLY-LIFE EXPOSURE TO CARCINOGENS (2005) available at http://www.epa.gov/raf/publications/pdfs/childrens_supplement_final.pdf.

89. See COMM. TO REVIEW EPA'S DRAFT IRIS ASSESSMENT OF FORMALDEHYDE, NAT'L RESEARCH COUNCIL, REVIEW OF THE ENVIRONMENTAL PROTECTION AGENCY'S DRAFT IRIS ASSESSMENT OF FORMALDEHYDE ch. 7 (2011).

90. *Id.* at 4.

91. *Id.* at 14, 152.

92. *Id.* at 152.

93. See Exec. Order No. 12,866, 58 Fed. Reg. 51,735 (Oct. 4, 1993).

benefits of regulation.⁹⁴ Moreover, with the black-boxing comes unexpected surprises once variables appear that are too important to ignore in future analyses. An example would be the potentially greater value of children and the need to monetize losses to children differently.⁹⁵

4. More General Executive Branch Directives Aimed at Scientific Integrity and Transparency

Highlights from some of the leading governmental directives further underscore just how pervasive this invisibility of methods is in contemporary regulation. In President Obama's memorandum and Office of Science and Technology Policy (OSTP) Director John Holdren's accompanying directive on government scientific integrity, the term "method" is never mentioned, and the concept of methods as a critical ingredient to ensuring both rigor and candor in the agencies' science discussions is wholly ignored.⁹⁶ Even in the agencies' own guidelines for improving scientific integrity, there is little movement toward making methods more visible or rigorous. Agencies instead focus their efforts on shoring up scientific misconduct programs (for fraud) or providing staff scientists with access to the press.⁹⁷ While these are important reforms, requiring agencies to identify and defend their methods of analysis seems at least as important for ensuring the integrity of regulatory science.⁹⁸

94. See, e.g., Wendy E. Wagner, *The Clean Air Interstate Rule's Regulatory Impact Analysis: Advocacy Dressed up as Policy Analysis*, in REFORMING REGULATORY IMPACT ANALYSIS 56, 56–57 (Richard Morgenstern et al. eds., 2009).

95. See, e.g., Sean H. Williams, *Statistical Children*, 30 YALE J. REG. (forthcoming 2013) (discussing how research on the value of children alters the way their losses should be calculated, which in turn could lead to considerable revamping of regulatory analyses for past and future regulations).

96. Memorandum on Scientific Integrity, 74 Fed. Reg. 10,671 (Mar. 11, 2009); Memorandum on Scientific Integrity from John P. Holdren, Assistant to the President for Sci. & Tech. and Dir. of the Office of Sci. & Tech. Policy, for the Heads of Executive Departments and Agencies pt. V (Dec. 17, 2010), <http://www.whitehouse.gov/sites/default/files/microsites/ostp/scientific-integrity-memo-12172010.pdf> [hereinafter Holdren, Scientific Integrity Memo].

97. See, e.g., WENDY E. WAGNER, SCIENCE IN THE ADMINISTRATIVE PROCESS: A STUDY OF AGENCY DECISIONMAKING APPROACHES 64–78 (draft Feb. 27, 2012), <http://www.acus.gov/wp-content/uploads/downloads/2012/02/COR-Science-Project-Report-2-27-12-CIRCULATED-TO-COMMITTEE.pdf>.

98. In its revised guidance for agency "risk assessments", for example, the Office of Management and Budget provides no reference to the need for a discussion of competing methods of inference, nor does OMB require the agencies to articulate how they synthesized the literature or developed their predictive models. See, e.g., Memorandum on Updated Principles for Risk Analysis from Susan Dudley, Admin., Office of Info. and Regulatory Affairs, Office of Mgmt. & Budget, and Sharon L. Hays, Assoc. Dir. & Deputy Dir. for Sci., Office of Sci. & Tech. Policy, for the Heads of Executive Departments and Agencies (Sept. 19, 2007), http://www.whitehouse.gov/sites/default/files/omb/assets/regulatory_matters_pdf/m07-24.pdf. One sentence of the twelve-page memo urges agencies to explicate

II. WHY METHODS ARE INVISIBLE

As the prior section makes clear, the invisibility of methods is not unique to the FDA, but can be seen across U.S. regulation. In these regulatory settings, moreover, policymakers, courts, and interest groups within the regulatory state all tend to focus their arguments and oversight primarily on regulatory outputs and/or overarching regulatory frameworks, rather than on the agencies' methods of reaching decisions.

This Part considers these institutional forces in some detail. Even though at its core the regulatory process is preoccupied with the rationality of decisions, what can be seen overall is the way in which regulatory processes have evolved in ways that make methods of inference invisible in regulatory debates.

A. Conventional Misunderstandings and Methodological Complacency Among Lawyers

One of most significant reasons that methods have been ignored is misguided impressions by the legal and policymaking communities that the methods of inference used by scientists lie beyond their expertise and thus are something to be avoided.⁹⁹ Thus, while the results of research are viewed as fair game for challenge, the way in which those results are generated are not. Such a misperception is not surprising. Issues of scientific inference do require expertise—a fact that anyone reading Part I would have felt acutely aware of—but to avoid these issues for this same reason is badly mistaken. Inferring from facts is part of the inherent rationality of a decision,¹⁰⁰ and the rationality of decision has been a constant theme in the history of the development of U.S. risk regulation.¹⁰¹ Indeed, the search for rationality in regulation has been the force behind legislative reform,¹⁰² executive oversight,¹⁰³ and judicial review.¹⁰⁴ Yet the search for rationality, at least within existing legal frameworks, stops quite short of considering methods of

their assumptions, judgments, inferences, and methods. *Id.* at 11. That is the only mention made of the need for agencies to extract their methods, including their methods of inference, and to make them explicit and accessible for comment.

99. Fisher et al., *supra* note 27, at 264.

100. *Id.* at 267–70.

101. See the analyses in COMM. ON RISK ASSESSMENT OF HAZARDOUS AIR POLLUTANTS, *supra* note 86, at ch. 2, and in NAT'L RESEARCH COUNCIL, SCIENCE AND DECISIONS: ADVANCING RISK ASSESSMENT ch. 2 (2009).

102. See *infra* Part II.B.3.

103. See THOMAS MCGARITY, REINVENTING RATIONALITY: THE ROLE OF REGULATORY ANALYSIS IN THE FEDERAL BUREAUCRACY (1991).

104. Patricia M. Wald, *Judicial Review in Midpassage: The Uneasy Partnership Between Courts and Agencies Plays On*, 32 TULSA L.J. 221, 227 (1996).

inferences.¹⁰⁵ While there are inherent challenges in understanding methods that lie on the other side of the disciplinary divide,¹⁰⁶ this does not mean that such methods are not important.

Second and relatedly, lawyers and policymakers have tended to presume risk assessment is a scientific method in and of itself, so that fundamental questions about how scientific analysis should be carried out are hidden from view.¹⁰⁷ Discourses about risk assessment have dominated risk regulation over the last two decades,¹⁰⁸ but there is little appreciation that these discourses have primarily been driven by concerns over administrative legitimacy rather than being explicit discourses about how best to infer conclusions from facts.¹⁰⁹ Indeed, in much of this discussion, science is largely understood as an input into the decisionmaking process that equates to the truth,¹¹⁰ a perception that obscures the complexity of methods and wrongly assumes that the methods are well established. In fact, in some policy circles there is an assumption that these terms are themselves methods. They are not, however, and as the National Research Council noted in 1994:

105. Prior commentary hovers around the importance of methods, but never lands on it squarely. As Judge Leventhal of the District of Columbia Circuit of the United States asked in 1976, “[w]hat does, and should, a reviewing court do when it considers a challenge to technical administrative decisionmaking?” *Ethyl Corp. v. EPA*, 541 F.2d 1,68 (D.C. Cir. 1976). That question led to a debate between him and Judge Bazelon over the nature of “hard look” review. *Id.* Edley has described Bazelon and Leventhal “as talking about two sides of the same coin, that coin being judicial activism motivated by a concern for sound governance.” CHRISTOPHER EDLEY, *ADMINISTRATIVE LAW: RETHINKING JUDICIAL CONTROL OF BUREAUCRACY* 227 (1990).

106. Fisher et al., *supra* note 27, at 277–79.

107. This concern is raised in reports such as COMM. ON HAZARDOUS RISK OF AIR POLLUTANTS, *supra* note 86, and NAT’L RESEARCH COUNCIL, *supra* note 101. See generally Wagner, *supra* note 25.

108. See NAT’L RESEARCH COUNCIL, *supra* note 101; see also COMM. TO REVIEW THE OMB RISK ASSESSMENT BULLETIN, NAT’L RESEARCH COUNCIL, *SCIENTIFIC REVIEW OF THE PROPOSED RISK ASSESSMENT BULLETIN FROM THE OFFICE OF MANAGEMENT AND BUDGET* (2007); 1 PRESIDENTIAL/CONGRESSIONAL COMM. ON RISK ASSESSMENT AND RISK MGMT., *FRAMEWORK FOR ENVIRONMENTAL HEALTH RISK MANAGEMENT* (1997).

109. See, e.g., Sheila Jasanoff, *The Songlines of Risk*, 8 ENVTL. VALUES 135, 137 (1999); FISHER, *supra* note 19, at ch. 3. Indeed, much of this discourse has been driven by a focus on outside-in accountability, where the focus has been upon controlling public administration, rather than inside-out accountability, which focuses on the methodology inherent within a discipline. Sidney Shapiro et al., *The Enlightenment of Administrative Law: Looking Inside the Agency for Legitimacy*, 47 WAKE FOREST L. REV. 463, 464–66 (2012). The outside-in vision of accountability can also be understood as promoting the rational-instrumental paradigm of administrative constitutionalism. FISHER, *supra* note 19, at ch. 3.

110. See, e.g., Wagner et al., *supra* note 17, at part II.A. While we did not elaborate on how this misunderstanding also obscures the importance of methods, it is clear from the examples we provide that policymakers tend to expect scientists and their processes to produce definitive answers to regulatory questions, without bogging the process down in discussions about alternative, plausible methods.

“Risk assessment is not a single, fixed *method* of analysis. Rather it is a systematic *approach* to organizing and analyzing scientific knowledge and information.”¹¹¹ As a result, critical decisions regarding methods of inference and for synthesizing the literature remain black-boxed while lobbyists insist that the agency employ “peer review,” utilize “risk assessment,” or ensure that their work is based on some form of aspirational “sound science.”

B. Reinforcing Institutional Incentives

Given the prevailing misperceptions of methods by the legal and policymaking communities, it is no wonder that institutional oversight processes tend to tolerate the invisibility of methods. What is perhaps more surprising is the tendency of some institutions to actively promote this invisibility in how they operate. These institutional problems are discussed next.

1. Science Advisory Committees

Science advisory committees provide perhaps the single most effective survival mechanism for an agency struggling to have its regulatory science accepted, and yet the science advisory process can work to keep methodological discussions out of public view. Rather than ensuring that methods are made more visible and subject to broader scrutiny, science advisory boards sometimes do nothing more than pass the scientific assessment from one black box process inside the agency to another, equally black box process involving advisory board review.

As a matter of orientation, science advisory boards used by the agencies are quite variable. Some science advisory boards are required by statutory mandate.¹¹² Others are employed by the agency on a program-wide or rule-specific basis.¹¹³ In most cases, agencies use the science advisors to review their assessments and proposed rules or standards.¹¹⁴ When they review

111. COMM. ON RISK ASSESSMENT OF HAZARDOUS AIR POLLUTANTS, *supra* note 86, at 4 (emphasis added).

112. *E.g.*, Clean Air Act § 109(d)(2)(A)–(B), 42 U.S.C. § 7409(d)(2)(A)–(B) (2006) (establishing a seven-member science advisory panel to consult with EPA on its review of the National Ambient Air Quality Standards under the Clean Air Act).

113. *See, e.g.*, GOV'T ACCOUNTABILITY OFFICE, GAO 08-440, CHEMICAL ASSESSMENTS: LOW PRODUCTIVITY AND NEW INTERAGENCY REVIEW PROCESS LIMIT THE USEFULNESS AND CREDIBILITY OF EPA'S INTEGRATED RISK INFORMATION SYSTEM 13 (2008), available at <http://www.gao.gov/assets/280/273184.pdf> (noting the participation of science advisory boards in the review of influential risk assessments under EPA's Integrated Risk Information System (IRIS) program).

114. For a richer discussion of how advisory boards do and should operate, see SHEILA JASANOFF, *THE FIFTH BRANCH: SCIENCE ADVISERS AS POLICYMAKERS* chs. 5–9 (1990); MARK R. POWELL, *SCIENCE AT EPA: INFORMATION IN THE REGULATORY PROCESS* (1999).

agency products, science advisory boards can provide assessments that are sometimes one-shot and in other cases are iterative.¹¹⁵

From the agency's perspective, science advisory boards offer the promise of insulating regulatory projects from broader attack against their scientific reliability, and thus agencies may seek out advisors to buffer them from these conflicts.¹¹⁶ As such, the agency tends to be less interested in stimulating open, frank academic debates among colleagues about the available methodological approaches and more interested in a yes-or-no consensus on the final regulatory decision. Agencies may even choose to design the advisory process to focus the group on their regulatory conclusions rather than methodological alternatives. When the task is framed in this way, the science advisors will likely comply with their assignment. Methods remain invisible and perhaps become even more deeply embedded in the layers of scientific review.

More recent process reforms of science advisory boards may exacerbate this black-boxing of methods even further. The White House Office of Management and Budget (OMB), for example, prescribes mandatory external peer review for influential rules, yet the OMB's detailed guidance says little to nothing about the need for expert review of the agencies' underlying methods.¹¹⁷ The apparent expectation is that an expert group will endorse or reject the agency's ultimate findings, rather than engage the agency in a dialogue about its choice and explication of methods and identify room for improvement. Most of the current regulatory commotion about the reform of the Federal Advisory Committee Act (FACA) and advisory board review is similarly focused on the selection of experts to serve on panels rather than ensuring that it is designed in ways that extract methods for larger expert and public review.¹¹⁸

115. See JASANOFF, *supra* note 114; POWELL, *supra* note 114. Generally, science advisory boards provide one-shot opinions on agency regulatory products. The EPA's advisory board (CASAC), which is consulted on EPA's review of various National Ambient Air Quality standards, is an exception to this rule. CASAC weighs in multiple times on each of EPA's scientific reports that supports a final revised standard. See, e.g., ENVTL. PROT. AGENCY, INTEGRATED REVIEW PLAN FOR THE NATIONAL AMBIENT AIR QUALITY STANDARDS FOR PARTICULATE MATTER 25, fig.4.1 (2008), available at http://www.epa.gov/ttnnaqs/standards/pm/data/2008_03_final_integrated_review_plan.pdf (identifying CASAC's involvement in EPA's review process).

116. See generally JASANOFF, *supra* note 114.

117. See OFFICE OF MGMT. & BUDGET, EXEC. OFFICE OF THE PRESIDENT, FINAL INFORMATION QUALITY BULLETIN FOR PEER REVIEW (2004), <http://www.whitehouse.gov/sites/default/files/omb/memoranda/fy2005/m05-03.pdf>.

118. See, e.g., BIPARTISAN POLICY CTR., IMPROVING THE USE OF SCIENCE IN REGULATORY POLICY 15-16, 41-42 (2009) (providing some of the most complete recommendations for the reform of science advisory boards, but dedicating very little of the proposals to the need for agencies to make their methods more accessible and explicit for this review).

2. Judicial Review

Since courts provide an external check on the agency to ensure that they have followed the requisite processes and explained their choices,¹¹⁹ the courts would seem the natural institutional check to ensure that methods have become visible in regulatory processes. The courts are where two of the most significant standards of review—the “arbitrary and capricious” and “substantial evidence” standards—are directly concerned with the rationality of decisionmaking.¹²⁰

Courts do require that agencies explain their decisions, and sometimes this explanation can lead to stays in a rulemaking until the court is satisfied.¹²¹ Courts also have power over agencies that softer political processes and interest group criticism lack. Thus, courts can provide a valuable lever to force agencies to make methods visible.

Courts, however, have managed to provide institutional oversight in a way that often ignores the need to force the agency to expose and explain its underlying methodological choices.¹²² Thus, for example, in applying the “arbitrary and capricious” standard of review,¹²³ the court tends to examine an appellant’s version of the “best available” facts and compare those facts to those used by the agency.¹²⁴ In analyzing facts, the focus tends to be upon issues of quantification and the use of assumptions rather than upon the choice of methods. Thus, writing for the Supreme Court in *Motor Vehicle*

119. See Administrative Procedure Act § 10, 5 U.S.C. § 706 (2006).

120. *Id.* Note, however, that in the risk regulation field many statutes set out their own standards of review. *E.g.*, Clean Air Act § 307(d)(9), 42 U.S.C. § 7607(d)(9) (2006); Toxic Substances Control Act § 19(c)(1)(B), 15 U.S.C. § 2618(c)(1)(B) (2006); Occupational Safety & Health Act §§ 6(f), 11(a), 29 U.S.C. §§ 655(f), 660(a) (2006). How these standards operate is complex. See Antonin Scalia & Frank Goodman, *Procedural Aspects of the Consumer Product Safety Act*, 20 UCLA L. REV. 899, 933–40 (1973); Elizabeth Fisher, *The Risks of Quantifying Justice: The Use of the Substantial Evidence Test in Judicial Review of OSHA Rule-Making*, in LAW AND UNCERTAINTY, RISKS AND LEGAL PROCESSES 293, 297–306 (Robert Baldwin ed., 1997) (explaining that “substantial evidence” was originally an adjudicative concept, but in the OSHA context came to be defined in scientific terms).

121. See generally Emily Hammond Meazell, *Super Deference, the Science Obsession, and Judicial Review of Agency Science*, 109 MICH. L. REV. 733, 738, 778–79 (2011) (discussing how courts seek explanations in their remands of rulemakings).

122. It should be added that these generalizations about judicial review are just that; there appears to be some, possibly considerable, variation around this mean. For a recent case that emphasizes and reviews methods as opposed to outputs, see *Coalition for Responsible Regulation, Inc. v. EPA*, 684 F.3d 102, 122–23 (2012) (per curiam). Whether this latest case signals a shift in the courts’ understanding and focus in their review of agency science remains an open question, however.

123. 5 U.S.C. § 706.

124. Thomas O. McGarity & Wendy E. Wagner, *Legal Aspects of the Regulatory Use of Environmental Modeling*, 10 ENVTL. L. REP. 10751, 10765–68 (2003) (discussing some of the litigation against EPA that makes this argument in the context of computational models).

Manufacturers' Ass'n v. State Farm Insurance Co., Justice Byron White stressed the need for an agency to "examine the relevant data and articulate a satisfactory explanation."¹²⁵ But while there is a resultant need to establish a "rational connection" between the facts and the choices made, this rational connection is not understood in terms of methods. Rather, it is more understood in terms of the factors taken into account; the plausibility of the decision depends on whether it is understood to be "a product of agency expertise."¹²⁶ Accordingly, the agency's methods are relevant only to the extent those methods *invisibly* support one body of evidence versus another.

Second, and reinforcing the judicial focus on "outputs," the courts seem to defer more heavily to agency outputs that have been reviewed and endorsed by science advisory panels.¹²⁷ By crediting this review as a plus in assessing the "arbitrariness" of the agency's finding, judicial review again reinforces the black-boxing of methods through advisory review. Methods become even more obscure and irrelevant to the test for rigorous regulatory science.

Third and finally, some have observed that the courts seem more deferential to technical "facts" than to candid discussions about competing assumptions and models.¹²⁸ Courts defer heavily to issues that are "on the frontiers of science,"¹²⁹ for example, yet when agencies concede that they faced policy-loaded choices in their methods, the courts sometimes scrutinize these contested decisions more rigorously. Furthermore, the focus is often upon agencies quantifying their decisions rather than explaining their methods.¹³⁰ As a result, agencies may rightly perceive that when they acknowledge their choices and decisions on method, they may find the courts more, rather than less, inclined to take a "hard look" and reverse decisions with which they disagree.

125. *Motor Vehicle Mfrs. Ass'n v. State Farm Mut. Auto. Ins. Co.*, 463 U.S. 29, 43 (1983).

126. *Id.*

127. This observation is at this point only an aesthetic observation based on some highly salient cases. *See, e.g.*, *Chlorine Chemistry Council v. EPA*, 206 F.3d 1286 (D.C. Cir. 2000); *see also infra* notes 174–177 and accompanying text.

128. *See, e.g.*, Richard J. Pierce, Jr., *Two Problems in Administrative Law: Political Polarity on the District of Columbia Circuit and Judicial Deterrence of Agency Rulemaking*, 1988 DUKE L.J. 300, 311 (bemoaning the willingness of the court to reverse agency policy choices when they are exposed).

129. *See, e.g.*, *Baltimore Gas & Elec. Co. v. Natural Res. Def. Council, Inc.*, 462 U.S. 87, 103 (1983) ("When examining this kind of scientific determination [i.e., one at the "frontiers of scientific knowledge"], as opposed to simple findings of fact, a reviewing court must generally be at its most deferential."); *Natural Res. Def. Council, Inc. v. EPA*, 902 F.2d 962, 968 (D.C. Cir. 1990) (deferring to agency decisions on frontiers of scientific knowledge), *vacated in part, appeal dismissed in part*, 921 F.2d 326 (D.C. Cir. 1991).

130. *See Indus. Union Dept., AFL-CIO v. Hodgson*, 499 F.2d 467, 475 (D.C. Cir. 1974).

Much of this situation is understandable and a natural response on the part of judges to the scientific content of decisions.¹³¹ Science is not in their area of expertise and the scope of review is limited.¹³² It is also perhaps due to the historical focus in administrative rulemaking upon the significance of a decisionmaker establishing a "rulemaking record." In the risk regulation context, courts have interpreted that administrative record as a purely factual one.¹³³ It is also the case that doctrines such as hard look review were adapted from more generic areas where issues of scientific method were less relevant.¹³⁴ Yet that does not detract from the fact that choices over method are not only inherent in the rationality of regulatory decisionmaking processes but also central to the question of how rational those processes are.

3. The Larger Regulatory Process

Agencies are not only bound by the Administrative Procedure Act (APA) and judicial review, but are also constrained by their authorizing statutes and a growing list of supplemental regulatory assessment requirements.¹³⁵ The invisibility of methods is thus not simply an agency creation or a result of judicial permissiveness, but in part, and in some cases in large part, attributable to legislative design. As a statutory matter, for example, Congress can lock the agency into a particular method, as it did with the FDA.¹³⁶ Simply by prescribing specific rulemaking requirements, methods become beside the point and fade into the background. While the FDA's drug program offers an illustrative example of this type of hard constraint, it is not alone. Under some authorizing statutes, for example, agencies are precluded from adopting a number of inferential approaches by statute.¹³⁷ While this more limited discretion in some cases appears to make the agency's life easier, the narrow delegations alter the agency's methodological choices

131. See Harold Leventhal, *Environmental Decisionmaking and the Role of the Courts*, 122 U. PA. L. REV. 509, 511-12 (1974).

132. See *Citizens to Preserve Overton Park v. Volpe*, 401 U.S. 402, 416 (1971).

133. See *Fisher*, *supra* note 120, at 299.

134. See *Motor Vehicle Mfrs. Ass'n v. State Farm Mut. Auto. Ins. Co.*, 463 U.S. 29, 40-46 (1983) (concerning the rescinding of a seat belt regulation); *Greater Boston Television Corp. v. FCC*, 444 F.2d 841, 850-52 (D.C. Cir. 1970).

135. For a recent inventory of some of the more significant requirements imposed on agencies, see CURTIS W. COPELAND, *REGULATORY ANALYSIS REQUIREMENTS: A REVIEW AND RECOMMENDATIONS FOR REFORM* 13-32 (2012), available at <http://www.acus.gov/sites/default/files/COR-Final-Reg-Analysis-Report-for-5-3-12-Mtg.pdf>.

136. See *supra* Part I.B.

137. The NAAQS reviews discussed *infra* Part III.B provide a good example of this more limited statutory delegation. See *Whitman v. Am. Trucking Ass'ns, Inc.*, 531 U.S. 457, 485 (2001) (interpreting the Clean Air Act to limit the discretion afforded the Administrator to consider economic consequences of health protective standards).

and how it can communicate them.¹³⁸ Congress can also lock into place methods for synthesizing the literature, as illustrated by the competing approaches to pesticide evaluation taken by the FWS and the EPA.¹³⁹

Additionally, both by statute and executive order, agencies must evaluate the impacts of their future rules on a range of targets such as small businesses,¹⁴⁰ the general economy,¹⁴¹ minority communities,¹⁴² etc.¹⁴³ Agencies must also be prepared to defend or subject technical information to an appeal process when their “facts” are challenged under the Data Quality Act.¹⁴⁴ Each of these accountability mechanisms demand “outputs”—with an expectation of considerable precision—that indicate whether the rule will impact small businesses or present an undue hardship on the economy relative to the regulatory benefits.¹⁴⁵ The programs do not even gesture toward the need for disclosure and explication of methods; rather they direct agencies to provide barometer-like readings on how the rules affect various, often conflicting features of American life.

The result of these regulatory accountability tools, again, is to focus the agencies on presenting answers without developing or explaining their methods. Competing methods for determining economic harms or considering the ways that rules might have adverse impacts on communities are wholly ignored. Such a blind spot is particularly ironic given the objective of these good government tools, which is to advance government accountability.

III. HOW TO CONSIDER METHODS

It is one thing to note that methods are invisible, but it is quite another to imagine how the regulatory discussions would look if methods were more centrally discussed in agency rulemakings. This is particularly difficult when there are fundamental intellectual challenges involved in evaluating methods in another discipline, particularly in circumstances of technical,

138. See Cary Coglianese & Gary Marchant, *Shifting Sands: The Limits of Science in Setting Risk Standards*, 152 U. PA. L. REV. 1255, 1266–74 (2004) (criticizing the Clean Air Act for an unrealistically narrow delegation).

139. See *supra* Part I.C.1.

140. See Regulatory Flexibility Act, 5 U.S.C. §§ 601–612 (2006).

141. See Exec. Order No. 12,866, 3 C.F.R. § 638 (1994), *reprinted as amended in* 5 U.S.C. § 601 (2006).

142. See Exec. Order No. 12,898, 3 C.F.R. § 895 (1995), *reprinted as amended in* 42 U.S.C. § 4321 (2006).

143. See generally COPELAND, *supra* note 135 (discussing the most taxing requirements).

144. Pub. L. No. 106-554, § 515, 114 Stat. 2763, 2763A-153 to -154 (2000).

145. One need only review the final sections of any proposed and final regulations for this checklist of applicable laws. See also COPELAND, *supra* note 135, at App. 2 (summarizing the requirements cited in 100 rules).

institutional, and political complexity.¹⁴⁶ Given the foundational role of methods in synthesizing scientific research to inform regulation, however, these challenges simply cannot be brushed aside or avoided. Methods are endemic in regulatory decisions and must be identified and confronted head-on.

There is no room here to enter into a detailed analysis of these issues. Our purpose here is to put the issue on the agenda for discussion. To begin the conversation, we offer some suggestions for what the regulatory discussion might look like *if* methods were made part of the regulatory discussions. The FDA experience with Vioxx demonstrates the importance of making methodological techniques accessible to policymakers, and the 2007 FDA legislation passed in response represents an important, corrective step that advances this recognition of the spectrum of inferential methods sketched in Figure 1, *supra*.¹⁴⁷ To provide additional illustrations of how agencies are making methods visible, we consider the Department of Interior's (DOI) use of probabilistic inference to justify its decision to classify polar bears as a threatened species under the Endangered Species Act (ESA). We also explore the EPA's staged approach to making methods visible in its regulation of air pollutants under the Clean Air Act. The section then closes with some preliminary thoughts about how the courts' approach to judicial review might be adjusted to provide greater encouragement for agencies to place methodological discussions centrally in their regulatory analyses and discussions.

Although there is considerable variation within different regulatory settings, some basic themes and principles emerge from this inventory of success stories. The examples in this section illustrate how the agency can explain its choice of inferential methods and how it can even identify how those methods suit the question at hand.

A. Department of Interior's Use of Probabilistic Inference

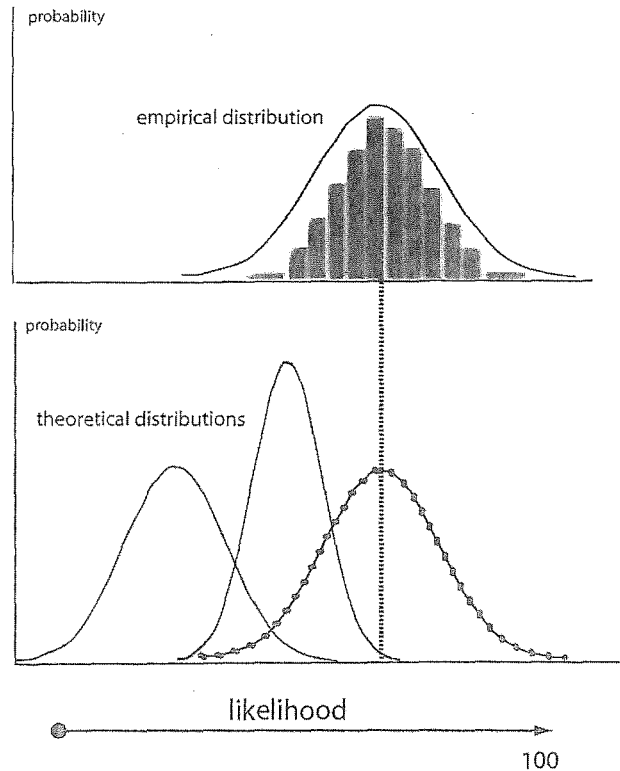
The DOI's decision to protect polar bears offers a concrete illustration of how agencies can articulate alternate methods of inference. Recall the discussion of one method of inference, *p*-value, in Part I *supra*. In their regulatory analysis of polar bears, the DOI employed a second type of

146. Fisher et al., *supra* note 27, at 279–82.

147. Note since the inferential methods generally run across entire programs or multiple programs, they may only need to be made explicit once and can be referenced after that. Thus the investment of energy and time in these meta-methodological decisions, which are also the most consequential, may not be substantial when parsed out over a number of regulatory projects.

inferential method—the use of likelihoods and Bayesian inference—described in Figure 4 below.¹⁴⁸

FIGURE 4. LOGIC BEHIND BAYESIAN INFERENCE AND LIKELIHOODS



Probabilistic inference rests on the premise that patterns in data of observed events can be estimated by probability distributions that can be mathematically manipulated. Unlike the method of p -values, which evaluates inferences based on whether empirical observations are consistent with a theoretical distribution, the methods of likelihoods and Bayesian inference share a common goal: *to estimate the theoretical distributions that best explain the observed events—even when the information about these events are heterogeneous.*

Bayesian inference assumes that if multiple events are related, then the probability distribution of each will affect that of all the others. Therefore, one should be able to predict how altering one event will influence the others.

If one has empirical data, one can evaluate competing theoretical distributions (bottom graph) based on a probabilistic measure called the likelihood, which can be regarded as the p -value's inverse. The likelihood measures the probability of a theoretical distribution, given the empirical (top graph). The most credible inference is the one based on the distribution with the highest likelihood (e.g., the rightmost distribution in the bottom graph).

148. For an overview of Bayesian thinking, see Pascual, *supra* note 25.

Under the ESA, the DOI must identify endangered and threatened species in need of protection.¹⁴⁹ The former are in danger of extinction, while the latter are likely to be so within the foreseeable future.¹⁵⁰ The DOI must make these determinations based on the best scientific and commercial data available to establish at least one of five factors: (1) present or threatened destruction, modification, or curtailment of habitat; (2) overuse for commercial, recreational, scientific or educational purposes; (3) disease or predation; (4) inadequacy of existing regulations; or (5) other factors affecting continued existence.¹⁵¹

Among the methods the DOI used to synthesize and integrate the scientific data on hand—thereby establishing the weight of evidence for the polar bear's threatened existence—was a computational model based on Bayesian inference. Given the method's underlying logic, the Bayesian model served as a transparent tool to integrate multiple strands of evidence into one cohesive system. The model (see Figure 5) consisted of three components:

Nodes represent the causes and intermediary effects influencing polar bear population. Note that the shaded boxes correspond to four of the five ESA factors listed above.

Arrows link these nodes in a causal chain of events.

Probability distributions determine how the state of one node affects the other nodes in the system.

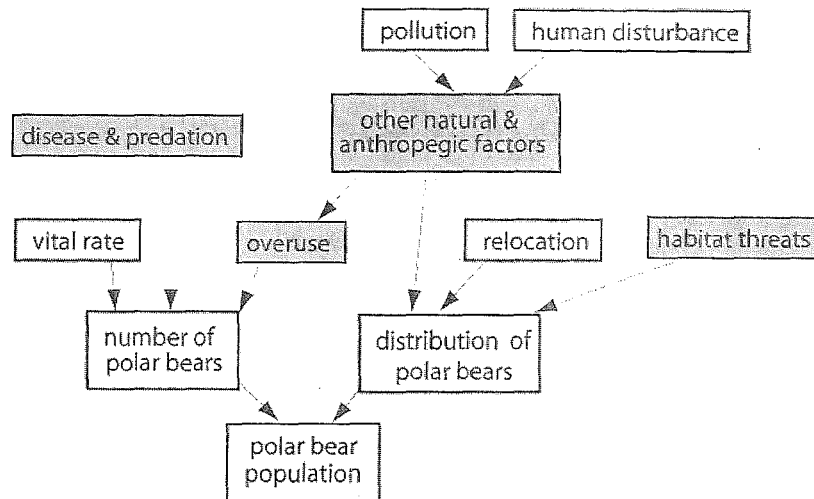
Taken together, these three components summarized the evidential narrative underlying the DOI's regulatory decision. This model then served as a formal means to integrate empirical data, expert judgment, model results, and other information within the DOI's assembled body of science. Some of these evidential components were individual studies in which evidence was evaluated by using *p*-values. The DOI's model therefore serves as an example of how multiple inferential methods may be used concurrently to evaluate the overall weight of evidence.

149. Endangered Species Act of 1973 § 4(a), 16 U.S.C. § 1533(a) (2006).

150. *Id.* § 1532(6), (20).

151. *Id.* § 1533(a)(1)(A)–(E).

FIGURE 5. BAYESIAN MODEL OF INTERIOR'S DECISION TO LIST POLAR BEAR AS A THREATENED SPECIES



DOI's model, based on Bayesian inference, has three components: *nodes* represent the system's major factors; *arrows* show the direction of causation; and *probability distributions* determine system behavior. The shaded nodes are those which DOI must consider under its statutory mandate.

Based partially on this model's results, the DOI in 2008 listed the polar bear as a threatened species.¹⁵² Shortly thereafter, numerous plaintiffs challenged the agency's decision in court.¹⁵³ One group claimed the animal merited greater protection as an endangered species, while the other claimed that the DOI should not have listed the animal at all. Both groups argued that the DOI had acted arbitrarily and capriciously in violation of the APA.¹⁵⁴

In its opinion, the district court acknowledged its narrow standard of review under the APA and the deference it owed to the agency, particularly for a regulatory decision requiring a high level of technical expertise.¹⁵⁵ But the court also emphasized its duty to hold the DOI to standards of rationality.¹⁵⁶ Following this path, the court scrutinized the agency's decision and

152. Endangered and Threatened Wildlife and Plants; Determination of Threatened Status for the Polar Bear (*Ursus maritimus*) Throughout Its Range, 73 Fed. Reg. 28,212, 28,274 (May 15, 2008) (to be codified at 50 C.F.R. pt. 17).

153. *In re Polar Bear Endangered Species Act Listing and § 4(d) Rule Litig.*, 794 F. Supp. 2d 65 (D.D.C. 2011), *aff'd*, Nos. 11-5219, 11-5221, 11-5222, 11-5223, 2013 WL 765059 (D.C. Cir., Mar. 1, 2013).

154. *Id.* at 78.

155. *Id.* at 80.

156. The court noted that it would remand the agency's rule if the agency "relied on factors which Congress has not intended it to consider, entirely failed to consider an important aspect of the problem, offered an explanation for its decision that runs counter to the evidence before the agency, or is so implausible that it could not be ascribed to a difference

upheld the DOI's decision. In the course of its affirmance, the court discussed the inferential challenges that lay before the DOI's decisionmakers, who needed

not only to evaluate a body of science that is both exceedingly complex and rapidly developing, but also to apply that science in a way that enabled them to make reasonable predictions about potential impacts over the next century to a species that spans international boundaries. [They] considered over 160,000 pages of documents and approximately 670,000 comment submissions from state and federal agencies, foreign governments, Alaska Native Tribes and tribal organizations, federal commissions, local governments, commercial and trade organizations, conservation organizations, nongovernmental organizations, and private citizens. [They] also consulted a number of impartial experts in a variety of fields, including climate scientists and polar bear biologists.¹⁵⁷

As *In re Polar Bear* reveals, interest groups did take issue with the DOI's model, but the developers of DOI's model provided measures of performance—based on model likelihoods, along with other techniques of model evaluation—to demonstrate the model's consistency with the evidence. By clearly delineating its understanding of the system of factors affecting polar bear populations and by declaring its judgment of the probable states of these factors, the DOI made its method visible in such a way that helped to focus discussion on the scientific merits and the appropriateness of its choice of inferences. The visible methods properly refocused the debate on the evidence and inferential choices rather than on the black-boxed result (*i.e.*, in favor or against classifying the polar bear as threatened) over which the litigants disagreed.¹⁵⁸ The DOI conceded that the model was only preliminary, that it was only one aspect of the evidence, and that it contained many uncertainties.¹⁵⁹ However, regardless of its results, the model helped confirm other evidence regarding the direction and magnitude of the effects of multiple stressors on polar bears.

The DOI's Bayesian model was just one of several strands of evidence used to justify the agency's decision. Our focus on this model is intended to emphasize how a variety of inferential methods can provide considerable evidentiary illumination in settings where the available evidence is broadly scattered and incomplete. As the court stated, the DOI did not rely on

in view or the product of agency expertise." *Id.* (quoting *Motor Vehicles Mfrs. Ass'n v. State Farm Mut. Auto. Ins. Co.*, 463 U.S. 29, 43 (1983)).

157. *Id.* at 68–69.

158. See Pascual, *supra* note 25.

159. *In re Polar Bear Litig.*, 794 F. Supp. 2d at 107.

existing data to establish statistically significant declines in polar bear populations. To justify its decision to list the species as threatened, the agency used models to predict significant *future* declines.¹⁶⁰ As we describe in Figure 5 and the accompanying text, for one of its models, the agency used Bayesian inference to communicate, in transparent and visible terms, the probabilistic underpinnings of its model, as well as the analytical foundations of the model.¹⁶¹ The joint plaintiffs did not contest the agency's choice of models, only the manner in which they were applied.¹⁶² The challenge from the joint plaintiffs therefore amounted to disagreements with the agency's judgments regarding the severity of climate change and of its effects on polar bear habitats and therefore, on their populations. Given the rational relationship between the agency's models and the reality they were purported to represent,¹⁶³ none of these disagreements compelled the court to abandon its deference to the agency in "an area characterized by scientific and technological uncertainty"¹⁶⁴

B. EPA's Synthesis of Scientific Information on Air Pollutants

Our next example takes a step back and highlights how in at least one program—the EPA's setting of air quality standards—the agency not only deploys multiple methods for inference and analysis, but has actually institutionalized a process to ensure that, to the greatest extent possible, methods are made visible.

As detailed in Section I, it is rarely the case that the weight of scientific evidence on human health and environmental risk will rest on the results of a single study. Typically, various strands of data, collected from multiple science investigations, will have to be woven together to reach an inference. And just as the multiple methods outlined in Figure 1 can be used to draw inferences from the data in a single study, so too can these multiple methods be used to integrate information from across multiple studies. Indeed, the model for polar bear populations described in the previous section was the DOI's attempt to use Bayesian inference to computationally integrate disparate information.

Under the Clean Air Act, the EPA must establish standards for ambient air concentrations of pollutants that "may reasonably be anticipated to endanger public health or welfare."¹⁶⁵ The EPA must revisit these so-called

160. *Id.* at 109.

161. For complete details, see STEVEN C. ARMSTRUP ET AL., U.S. DEP'T OF INTERIOR, FORECASTING THE RANGE-WIDE STATUS OF POLAR BEARS AT SELECTED TIMES IN THE 21ST CENTURY 12–19 (2007).

162. *In re Polar Bear Litig.*, 794 F. Supp. 2d at 108 n.51.

163. *Id.*

164. *Id.* at 108 (citations and internal quotation marks omitted).

165. Clean Air Act § 108(a)(1)(A), 42 U.S.C. § 7408(a)(1)(A) (2006).

National Ambient Air Quality Standards (NAAQS) at least once every five years.¹⁶⁶ In doing so, it must consult with an independent, scientific committee—the Clean Air Scientific Advisory Committee (CASAC).¹⁶⁷ Although not statutorily required to do so, as a matter of course, the EPA prepares reports to document the methods and the rationale it uses to integrate the scientific information culled to justify the agency's air regulations.¹⁶⁸ In these reports, the EPA divides the analytical project into distinct steps that allow it to better articulate its methods of analysis.¹⁶⁹ First, the EPA crystallizes the policy questions. In a second report, the EPA then assembles and synthesizes the relevant scientific literature that has bearing on those policy questions. In a third report, the EPA applies a variety of alternative models to the available scientific literature to reach predictions about air quality and public health impacts. The EPA concludes the exercise with a report that explains for sophisticated policymakers the key methodological steps that it used in the analysis, the range of conclusions and uncertainties surrounding different possible standards, and highlights research questions for the future.

Two recent legal challenges to the EPA's NAAQS—one for nitrogen dioxide (NO₂) and another for particulate matter (PM_{2.5})—illuminate the EPA's institutionalized process for transparently integrating various methods into its scientific inferences. In 2005, the EPA began to consider epidemiological and clinical evidence suggesting that respiratory illnesses were occurring at lower NO₂ concentrations and at shorter durations of exposure than had previously been thought. The EPA published a call for information in the *Federal Register*,¹⁷⁰ and in 2007, issued a research plan in which it discussed the major science issues to be addressed, the methods it would use, and its intent to present its results before the CASAC.¹⁷¹ In 2008, the agency published its assessment of the science.¹⁷² The CASAC agreed with

166. *Id.* § 7409(d)(1).

167. *See id.* § 7409(d)(2)(A)–(B).

168. *See Process of Reviewing the National Ambient Air Quality Standards*, U.S. EPA, <http://www.epa.gov/ttnnaqs/review.html> (last updated Dec. 10, 2012). Each criteria air pollutant is listed on the left bar. By clicking the pollutant, one can view the various reports that have been issued.

169. *See, e.g., NAAQS PROCESS REVIEW WORKGROUP, EPA, REVIEW OF THE PROCESS FOR SETTING NATIONAL AMBIENT AIR QUALITY STANDARDS E-1 (2006) available at* http://www.epa.gov/ttnnaqs/pdfs/naaqs_process_report_march2006.pdf.

170. Call for Information, 70 Fed. Reg. 73,236 (Dec. 9, 2005).

171. EPA posted all of these planning documents on its website at *Nitrogen Dioxide (NO₂) Primary Standards—Documents from Review Completed in 2010—Planning Documents*, U.S. EPA, http://www.epa.gov/ttn/naaqs/standards/nox/s_nox_cr_pd.html (last updated Dec. 10, 2012).

172. The EPA's various drafts of its integrated science assessment are posted at *Nitrogen Dioxide (NO₂) Primary Standards—Documents from Review Completed in 2010—Integrated Science Assessments*, U.S. EPA, http://www.epa.gov/ttn/naaqs/standards/nox/s_nox_cr_

the EPA's assessment. The EPA proposed new NAAQS in 2009 and published its final rule in 2010.¹⁷³ It bears highlighting that this entire process involved notification in the *Federal Register*, multiple rounds of public comment, and publication of the various science documents on the web.

In *American Petroleum Institute v. EPA*,¹⁷⁴ the petitioners claimed that this NO₂ standard was arbitrary and capricious under the APA because the EPA misconstrued the scientific evidence and relied on non-peer reviewed materials. The court concluded that on every alleged breach of scientific judgment—i.e., that EPA ignored countervailing evidence regarding NO₂'s effect on respiratory illness; that it misused an epidemiology study; and that it based the projections of the rule's benefits on faulty assumptions—the agency provided a reasoned defense of its inferences. The court stated that “[p]erhaps the [petitioners themselves] should have had [their] brief peer-reviewed.”¹⁷⁵

On the other hand, in *American Farm Bureau v. EPA*,¹⁷⁶ the court ruled that the EPA *was* arbitrary and capricious in failing to explain adequately why it did not consider certain scientific evidence that supported a more stringent NAAQS for fine particulate matter.¹⁷⁷ Just as it did when developing the NO₂ standard in the preceding paragraph, the EPA issued the standard for PM_{2.5} after a process that entailed both the crafting of reports to summarize the science behind the proposed NAAQS for PM_{2.5}, as well as consultations with the CASAC.¹⁷⁸ The EPA's political management, however, rejected the ultimate recommendations emerging from the five-year scientific process and decided that the high costs of the standard,

isi.html (last updated Dec. 10, 2012), and its drafts of the risk and policy assessments are posted at *Nitrogen Dioxide (NO₂) Primary Standards—Documents from Review Completed in 2010—Risk and Exposure Assessments*, U.S. EPA, http://www.epa.gov/ttn/naaqs/standards/nox/s_nox_cr_rea.html (last updated Dec. 10, 2012).

173. Primary National Ambient Air Quality Standard for Nitrogen Dioxide, 74 Fed. Reg. 34,404 (proposed July 15, 2009); Primary National Ambient Air Quality Standard for Nitrogen Dioxide, 75 Fed. Reg. 6,474 (Feb. 9, 2010) (to be codified at 40 C.F.R. pts. 50, 58).

174. *Am. Petroleum Inst. & Util. Air Regulatory Grp. v. EPA*, 684 F.3d 1342 (D.C. Cir. 2012).

175. *Id.* at 1348.

176. *Am. Farm Bureau Fed'n v. EPA*, 559 F.3d 512 (D.C. Cir. 2009).

177. *Id.* at 522. Fine particulate matter, or PM_{2.5}, consists primarily of soot particles with diameter less than 2.5 nanometers. This air pollutant is linked to higher levels of mortality and morbidity. See *Particulate Matter*, EPA, <http://www.epa.gov/air/particlepollution/> (last updated Jan. 23, 2013).

178. For an inventory of the various documents and drafts on EPA's particulate NAAQS review, as well as the comments and EPA's responses, see *Particulate Matter (PM) Standards*, EPA, http://www.epa.gov/ttn/naaqs/standards/pm/s_pm_index.html (last updated Dec. 20, 2012).

coupled with evidentiary uncertainties, favored a less stringent standard.¹⁷⁹ In reversing this final standard, the court compared it with the range of plausible standards developed through the EPA's institutionalized scientific process, which provided a transparent and rigorously reviewed discussion of methods and accompanying uncertainties.¹⁸⁰ The court concluded that when set against this robust scientific record, the EPA's different final standard lacked a rational basis.¹⁸¹

Both the polar bear and the EPA's PM_{2.5} case illustrate that, once methods are made visible, the agency's ultimate conclusion rests on a more solid foundation. Rather than being vulnerable to tangential nitpicking, the articulation of methods provides support for standards that are the result of multiple inferences, none of which can be firmly grounded in existing scientific knowledge. This is not to say that decisions will not be criticized or that they will not be subject to challenge. Rather, the focus of criticisms and challenges will be on the methods that are fundamental to the decision. Again, to draw on our legal analogy, within legal scholarship we understand the distinction between legitimate and irrelevant grounds of criticism in relation to judgments. In the case of the EPA's rules, moreover, the ability of the EPA to point to this type of careful explication of its methods is a relatively new and welcome innovation. The EPA only recently revised its NAAQS process, and prior to this renovation, the EPA's NAAQS process suffered from the same invisibility of methods—and the accompanying litigation, political controversy, and related strife—as most other agency protective standards. The NAAQS process thus offers a valuable before-and-after portrait of the institutional attributes of visible methods.

C. Encouraging Visible Methods by Adjusting Judicial Review

Because the EPA's revised NAAQS process in large part resulted from consistent, strong pressure from the courts and litigation, we close this Section by considering ways that judicial review and court directives might be used to affect a gradual shift away from existing incentives that tend to reward the invisibility of methods. The role of the courts is important. Indeed, judicial review of an agency decision is essentially the tail that wags the regulatory dog; the EPA's treatment of scientific evidence occurs in expectation that the evidence will be the subject of considerable scrutiny in

179. Juliet Eilperin, *Proposed Standards for Air Quality Criticized*, WASH. POST, Dec. 21, 2005, at A10 (describing the Bush Administration EPA's decision to reject a more stringent particulate standard despite strong scientific evidence, including CASAC endorsement, in its favor).

180. See *Am. Farm Bureau Fed'n*, 559 F.3d at 520–24.

181. *Id.*

the courtroom.¹⁸² The courts therefore exert considerable influence in compelling agencies to make their inferential methods visible and to ensure that these visible methods conform to the APA's mandate for rational governance.

Given the ad hoc and context-specific nature of their task, it has not been easy for courts to evaluate whether the inferential link between data and inference tips the scale to a point where the weight of evidence supports an agency's claim of a rational basis.¹⁸³ Instead, given these challenges, it has been easier for courts to effectively abrogate their responsibilities under the APA and simply defer to agencies.¹⁸⁴ As the court in *In re Polar Bear* noted, however, while judicial review of regulatory science is narrow, when conducting this review, courts should be able to understand what the agency's methods of inference were and why the agency deemed them appropriate to the task at hand.¹⁸⁵ Without such a basic explanation, the legal community lacks a principled basis for evaluating regulatory science.

Under our proposal, the courts should compel the visibility of methods by ensuring that the agencies provide answers to two deceptively simple questions: (1) *Have the agency's methods of inference been identified?* and (2) *Does the agency explain how its methods are appropriate to the information on hand and how they support the ultimate inference used by the agency?* Unless an agency can respond to both these questions in the affirmative, then the agency's science-based decisions should risk reversal or remand by the courts. Courts already require an "explanation" of the agency's choices.¹⁸⁶ Our proposal requires that the agency describe how it drew its inferences and identify the specific assumptions it made in the course of assembling the scientific evidence.

For example, because the FDA persisted in using *p*-values and randomized, clinical trials to evaluate Vioxx even when evidence indicated variable response to Vioxx within a heterogeneous population, the answer to both questions would have been a resounding "no."¹⁸⁷ Because the DOI explained how it used and decided upon the underlying probabilities and how it evaluated its model to estimate polar bear population, the court could answer both questions in the affirmative.¹⁸⁸ The EPA's detailed explication of its

182. The statement is based on the primary author's personal observation based on almost twenty years of experience working on regulatory science.

183. See Pascual, *supra* note 25; Christopher H. Schroeder & Robert L. Glicksman, *Chevron, State Farm, and EPA in the Courts of Appeals During the 1990s*, 31 ELR 10371 (Apr. 2001).

184. *Schroeder & Glicksman, supra* note 183.

185. See 794 F. Supp. 2d 95 (D.D.C. 2011).

186. See, e.g., Meazell, *supra* note 121, at 738, 778–79 (discussing how the courts tend to require explanations from the agency regarding its science-based decisions).

187. See *supra* Part I.B.

188. See *supra*, Part III.A.

methods of analysis, embodied in a succession of reports each of which was subject to expert review and public comment, provided a solid bottom to a final nitrogen dioxide standard that emerged from that process. This same institutionalized process also spotlighted the lack of support for a PM_{2.5} standard that diverged from agency staff and peer reviewer recommendations.

Under this proposal, once the agency believes that it can stand by its methods and respond affirmatively to these two questions, the burden should shift to those who would challenge the agency's scientific conclusions. Challengers currently gain credit by launching critiques against the agency's findings that highlight missed studies or data, flawed assumptions, or unaccounted for differences between the agency's model and the real world.¹⁸⁹ In most cases, participants are not required to show how their preferred variables, studies, or assumptions serve as a definitive improvement over the agency's version. The courts should require challengers to demonstrate that their methods mark a decided improvement over the agencies' approach.

Rather than positioning the challengers in a way that encourages sandbagging and second-guessing, under this proposal the challengers serve as constructive contributors to the development of more robust methods. Moreover, since challengers must give the agency notice of their criticisms in the comments, the agency will have the benefit of alternative methods during notice and comment and can account for innovations and other salient arguments earlier in the process.

CONCLUSION

In today's political climate, the terms "sound science" and "junk science" are bandied about by both ends of the political spectrum to advance their own regulatory agenda. Lost in the cacophony over scientific evidence is the cautionary statement offered by the scientist Werner Heisenberg: "[W]hat we observe is not nature itself, but nature exposed to our method of questioning."¹⁹⁰ We do not argue for some form of positivism or for a naïve search for objectivity. Rather, we contend that agencies make their "methods of questioning" visible so that courts can determine "the

189. See, e.g., 1 RICHARD J. PIERCE, JR., ADMINISTRATIVE LAW TREATISE § 7.4, at 594 (5th ed. 2010) ("If a comment criticizes in detail some characteristic of the agency's proposed rule, and the agency retains that characteristic in the final rule without including in its statement of basis and purpose a relatively detailed response to that criticism, a reviewing court is likely to hold the rule unlawful . . .").

190. WERNER HEISENBERG, PHYSICS AND PHILOSOPHY: THE REVOLUTION IN MODERN SCIENCE 26 (1958).

crucial difference between what is well or badly constructed, well or badly composed.”¹⁹¹

Making methods visible is only the first of several analytical steps—the methods also must be robust and appropriate for the analysis at hand.¹⁹² Yet until this first step is taken, we can only stab in the dark and imagine what inferential methods and analytical assumptions the agency might have employed. As this article details, agencies can do better. Methods can be made visible. And when they are, the discussion lays the essential foundation for ensuring a more productive approach to using science for regulation.

191. Latour, *supra* note 70, at 474 (emphasis omitted)(arguing for a ‘compositionist manifesto’ in which scholarship moves beyond simple critique). We recognize that culture will shape the answer to this question. See Sheila Jasanoff, *A New Climate For Society*, 27 THEORY, CULTURE AND SOCIETY 233 (2010).

192. We are not blind to the fact that a “cook-book” approach to inference would lead to more definite legal outcomes. Courts could insist that statistical significance—based on the method of *p*-values, as applied to the results of controlled, randomized trials—is the best and only way to substantiate the weight of evidence. To do so would be a legal fiction that contravenes scientific thinking. See J. Worrall, *Causality in Medicine: Getting Back to the Hill Top*, 53 PREVENTIVE MED. 235 (2011). To do so would perpetuate an irrational approach to rationality. We are similarly aware that a pluralistic approach to inferential methods has profound and far-reaching consequences on how the legal community understands causation. See R. Scheines, *Causation, Truth, and the Law*, 73 BROOK. L. REV. 959 (2007). We look forward to addressing these complications in future work.